

STIC-Biotech/ChemLib

75771

From: Rao, Manjunath N.
Sent: Monday, September 16, 2002 9:27 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/995,587

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10C 01
Phone: 306-5681

Date: 9-16-02

Please search the following as soon as possible for application with serial number **09/995,587**

SEQ ID NO: 2 and 10 against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

SEQ ID NO:1 and 11, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If y u have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Crystal Mall 1, USPTO.

RECEIVED
SEP 16 2002
STIC/BIOTECH DIVISION

No 100% or
65-100% matches
for both DNA & protein
No matches for 15.aa
fragment was found.

Point of Contact
P. Sheppard

Searcher: Telephone number: (703) 308-4499
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 9/27/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____


```

QY 320 FGYNSTAV-----SQWGSGLVNSDNLSTOLFRTVDTSDNNTN---HOKIASATL 367
Db 152 FKDNKLLANDPILKNOSEWGSATLSDGQVRLFTYSRQYDPNPNLYKQTLSTAQI 211
QY 368 YLTD-NNGNVSLAQVRNDYIYFE-CDGYYQTYDQWATKNGADNIARDARHVEDNGD 425
Db 212 NVSQPDDTKLKDGVEDLKSIDYDGGDKIYQNVQ--SVGVDMNDHTRDPHYVED-QGH 268
QY 426 RYLVEASTGLE-NYQGEDQTYNNLNGYGGDAFNKLSFRILSNDIDKSRATWANAAGI 484
Db 269 KYIIFEAATGTETGYQGEDSTQNPAIYGGNKKFFTEBQOQLLQSPK-KKGAELANGALGI 327
QY 485 LKLNKDEKNPKVAELYSPLISAPVSDIERPNVVLGNKYLYFAATRLNRGNSDDAWMN 544
Db 328 VELNDY---TLKNVMPPLIASNLVTDTERANVFKMGLWILFTSTRGSKVTVD----- 379
QY 545 ANYAVG-DNVAMVGYVADSLTSGYKPLNDGVLVLTASVPANNRTATSYIYVAVPAGKDDQ 603
Db 380 ---AIGDDDIYMLGYVSTSLTGPYKPLNGTGLVHLQDLDRDDVWTYAHFAIP-OGKGN 435
QY 604 VLVTSTYTNRNGVAGKGM-----DSTWAPSFLLQINPDNTTTLAKWTNQGWIMWDSSE 658
Db 436 VVVSMTNR-----GLFPDHKSTFAPSELLNKGSKTSYVKNGLILEQGGITVDPTND 488

RESULT 2
Q97181
ID Q97181 PRELIMINARY; PRT; 489 AA.
AC Q97181
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE LEVANSUCRASE.
GN CAC1772.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Neelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT *Genome sequence and comparative analysis of the solvent-producing
RL bacterium Clostridium acetobutylicum.*;
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007686; AAK79737.1;
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
KW Complete proteome.
SQ SEQUENCE 489 AA; 54553 MW; 9002CB364F1D3CD8 CRC64;

```

```

Query Match
Best Local Similarity 17.1%; Score 707; DB 16; Length 489;
Matches 183; Conservative 88; Mismatches 173; Indels 48; Gaps 20;

QY 188 LTDDQIKALNKNFSKAASKTQMTYNDFOKRIADTLIKQGRYVFPFKASEIKNMPAAT 247
Db 21 ITQOTFASTNDMT-KETGYVSHITRYNMSKI--PHEQNDLKFKVPQFNASTLKNIASAK 77
QY 248 TKDAQWTPIELDVWDSWPQDVRTGOVANNQYQLVIAAMGIP-NQNDNHVILLYNKG 306
Db 78 GYDKNGNLID-LDVWDSWPLQN-CGDTVANYHGYHVFALAGDPKQDDTSYMPYQKIG 135
QY 307 DNELSHWKNVGPFGYNSTAV-----SQWGSGLVNSDNLSTOLFRTVDTSDNNTN--- 355
Db 136 ENSIDSKNAGKVFKDSKYVANDPYLKYTOEWGSGATLSDGQVRLFTYDFSGVAKDG 195
QY 356 NT---NHQKIASATLYLTNDNGN-VSLAQVRNDYIYFE-CDGYYQTYDQWATNK--GAD 409

```

```

Db 196 GTDASNOVITTTQVNLSDPNSNTINISVSDHKSVDGCGNGRTIYNQVQFIDEGKWSGD 255
QY 410 NIAMRADHVEDGANGRYLVEASTGL-ENYQGEDQIYNWLYNGGDAFNKLSFRILSN 468
Db 256 NHFLRDPHYVED-NGRKYLVEANTGTNDGYOGDTSLNKAIFYGRSQSFFKTEKQDLID 314
QY 469 DDIKSRATWANAAGIATLKNKDEKNPKVAELYSPLISAPVSDIERPNVVLGNKYLYF 528
Db 315 TNKKHDASLANGALGIELNNDYTLKKEMK---PLIASNTVTDIERANVFKMGRWYLF 371
QY 529 AATRLNRGNSDDAWMNANYAVGDNVAMVGYVADSLTSGYKPLNDGVLVLTASVPANNRTA 588
Db 372 TD---SRGSK-----MTINGISSKDIYMLGFSNLSLTGPYKPLNGTGLVNLNDPTDLTF 424
QY 589 TSYIYVAVP-VAGKDDQVLTSTYTNRNGVAGKGM-----DSTWAPSFLLQINPDNTTTLA 643
Db 425 TYSHFAVPQTNGKN--VVITSYITNR-----GMYSDDHSSFAPSELLNKGKTSVSN 476
QY 644 KMTNOGDWIDDD 655
Db 477 SILQQQLTIDN 488

```

```

RESULT 3
ID O82854 PRELIMINARY; PRT; 487 AA.
AC O82854
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE BETA-FRUCTOFURANOSIDASE.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V230;
RA Tsuchi K.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kurimoto M.;
RT "Cloning and sequencing of beta-fructofuranosidase gene from Bacillus
RL sp. V230.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB010272; BAA32083.1;
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
SQ SEQUENCE 487 AA; 53412 MW; 7F52A47921824AE3 CRC64;

```

```

Query Match
Best Local Similarity 14.9%; Score 613.5; DB 2; Length 487;
Matches 168; Conservative 81; Mismatches 158; Indels 87; Gaps 21;

QY 199 MNFSAKASGT-----QMTYNDFOK-----IADTLIKQDG-- 228
Db 1 MNFKRLAKAAAVTFTAILVAGDGPPIFAQOMSGDKEDYGFHITRAD-MLKIPQQ 59
QY 229 ---RTVFPFKASEIKNMPAATTKDAQTNTTEPLDVWDSWPQDVRTGOVANNQYOLVI 285
Db 60 NSPOFKVPQFQNSAKNIDSAGYKSGNKLID-LDVWDSWPLQN-DCTAANYHGHIVS 117
QY 286 AMMGIP-NQNDNHVILLYNKGYNELSHWKNVGPFGYNSTAV-----SQWGSGLVNSDNLSTOLFRTVDTSDN- 335
Db 118 ALADPNKSDOTPLHLFYQKVGDTSIDSWKNAGRVFEDMDKFPNDPYLKYTOEWGSGA 177
QY 336 VLNSDNLSTOLFRTVDTSDN-----NTNHQKIASATLYLTID-NNGNVSLAQVRNDYIYF 388
Db 178 TLTQDQVRLFTY--DYSGNPDDGGTGAGNOIISTAQVNLSDPDAATLKVDGVSDDHKS 235

```

QY 389 E-GDGYTYQYDQWKATK--GADNIAMRDHVIEDNGDRYLVEASTG--LENYQGEDQ 444
 Db 236 DGGDGTVTQNIQQFIDECKWISGDNHTLRDPHYVED-KGHYLYVEANTGTDGYQGDQS 294
 QY 445 IYNWNLNMGDDAFNKSIFRILSNDNDIKSRATWANAAGILKLNKDKNPKVAELYSPLI 504
 Db 295 FNNKAYYGGSDVFFONEKNKLLQSPK-KQIASLANGALGIVELADDY---TVKSVMKPLV 350
 QY 505 SAPMVSDEIERPNVVKLGKLYLEAATRLNRGSDNDAMNANAVGD-NYAMGYVYVADSL 563
 Db 351 ASNTVADEVERANIFKMNKWLFTDSRGSKWTSN-----GINDKOVYMLGPGGDSL 402
 QY 564 TGSYKPLNDSGVLTASVPANWRATYSYAVPVAGKDDQVLTYSYMTNRN----- 614
 Db 403 NGPHNPINETGLVLMNLDPADLTHYSHCGIP-HPEGNNVLTYSMTNRGFEYPEHSHL 461
 QY 615 ----GVAGKGMDS 624
 Db 462 RDLGVNIRKSGSDTS 475

RESULT 4
 Q97179 PRELIMINARY; PRT; 428 AA.
 ID Q97179
 AC Q97179
 DT 01-OCT-2001 (TremBLrel. 18, Created)
 DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
 DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE LEVANSUCRASE.
 GN CAC174.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Neolling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R.L., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum."
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AE007686; AAK79739.1;
 DR InterPro; IPR003469; Glyco_hydro_68.
 DR Pfam; PF02435; Glyco_hydro_68; 1.
 KW Complete proteome.
 SQ SEQUENCE 428 AA; 48395 MW; 9670B154B178E23E CRC64;

Query Match 8.6%; Score 356.5; DB 16; Length 428;
 Best Local Similarity 27.2%; Pred. No. 5.4e-09;
 Matches 125; Conservative 70; Mismatches 176; Indels 89; Gaps 22;

QY 209 TQMTYNDFOKIADTLIKDGRYTPF-----FKASEIKNNPAATTKDAQTNTI 256
 Db 3 TRKTY-----KMISSLWVLAITLTPFILRHNTGYTSINSRQAQKFC--TKENTAPNI 56
 QY 257 EP-----LDVWDSWPFQDVRGTGOVANNVGYLVITAMGIPN--QNDNHIYLLNKKYG 306
 Db 57 NPNEKLTAPNLWVWDTPLYK-KDGLAVVNGYKVPALTASRVGNKRDHVGAGISYFC 115
 QY 307 DNELSHKKNVPIFGYNSTAVSQWGSVLSNDSNLIQIFYTRVDSN--NTNHOKIAS 364
 Db 116 STDGENVYKGLAYNVDELALGSRWAGSAILDENGWQFFYTATGRKEAVRTFEGRLYK 175
 QY 365 ATLYLTNDNNGNVSLAQVRNDYIVFEGDGYTYQYDQWKATNKGADNIAMRDHVI 424
 Db 176 TKFSINDKGVGHTNCSKHQVILEPDGVYTYQYDQWKATNKGADNIAMRDHVI 491

QY 425 DR-YLVEASTG--LENYQGE---DQIY--NWLNYGGDDAFNKSIFRILSNDNDIKSRAT 476
 Db 232 KKDYLIFEGNGKGIEMKMPENIGDKLFRKNHJAPRGVENFN----- 273
 QY 477 WANAAGILKLNKDKNPKVAELYSPLISAPMVSDEIERPNVVKLGKLYLEAATRLNRG 536
 Db 274 -GNVGIATAO--NKDLTRFK---LLPPLLEAVGVNQQLERQIVMKKNKYYLTISH--- 324
 QY 537 SDDANWNNAYAVGDN--VAMGYVADSLTGSYKPLNDSGVLTASVPANWRATYSY 594
 Db 325 -----KTYAGGLNGVDGLYGCNSRSNKPLNGNLVITN--PTNDPYQYYSWL 375
 QY 595 VPVAGKDDQVLTYSYMTNRNGVAGKGMDSWAPSELLQIN 634
 Db 376 --VSGHDVLSFINEY--HFGQLRYG--GTEAPT--LQIS 407

RESULT 5
 Q9EVD6 PRELIMINARY; PRT; 630 AA.
 ID Q9EVD6
 AC Q9EVD6
 DT 01-MAR-2001 (TremBLrel. 16, Created)
 DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE FRUCTOSYLTRANSFERASE.
 GN FTF.
 OS Actinomycetes naeslundii.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.
 OX NCBI_TaxID=1655;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 12104, WVU45;
 RA Bergeron L.J., Morou-Bermudez E., Burne R.A.;
 RT "Characterization of the fructosyltransferase gene of Actinomycetes
 naeslundii."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF228592; AAG09737.1;
 DR InterPro; IPR003469; Glyco_hydro_68.
 DR Pfam; PF02435; Glyco_hydro_68; 1.
 KW Transferase.
 SQ SEQUENCE 630 AA; 68256 MW; CDCA04184D6D7D2 CRC64;

Query Match 7.7%; Score 318; DB 2; Length 630;
 Best Local Similarity 22.7%; Pred. No. 4.6e-07;
 Matches 158; Conservative 93; Mismatches 236; Indels 210; Gaps 32;

QY 50 IAVKSVTLGSGOVSAAASDT---TIRTSANANSASSANTONSQVASSAAITSSSSAA 106
 Db 19 LAGLIVAGTGTALADETPSPSSSATAEASPEAGADQKNDQSPADQQAQTPTDQG 78
 QY 107 SLNNTSKAAQENTNTAKNDDTQKAAPANESSKAPNPAVNVNDSSAAKND--QQSSKKN 165
 Db 79 QAQAD-----QOAGGAAQADNGFKADNPQGANATKHTGAAGHVEEN 122
 QY 166 TTAKLNK-DA---ENVVKKAGIDPNLSLTDQIKALNKNMFSKAASGCTQNTYDFOKIAD 221
 Db 123 YTAKTATRAMQIQRYVFPQG-----AVRAPTRL----- 151
 QY 222 TLIKODGRYTVPFKASEIKNNPAATTKDAQNTTPELDVWDSWPFQDVRGTGOVANNWY 281
 Db 152 -----PEQLTMP-----EISNGFPATSEVW-----VMDTWLTDEAAHQI--SYNGW 192
 QY 282 QLVIAMMGIPNQ-----NDNHIY-----LLYKNYGDNELS-----HWKNVGPFGNS 324
 Db 193 EIAFSLVADRHHAGYTFDRHTRHARLGEFFYRKAGTQTSADGAGSSNGWYIGHVFPDAS 252
 QY 325 TAVSO-----EWSGSVLSNDSNLIQIFYTRVDSN--DNNTNHOKI----- 362
 Db 253 ARSSRTVHAQTENWGSARLMKIRMEKIRFYTSPSTTAPPRTGTTHQVTHATSKPYDPR 312
 QY 363 ---ASATLYLTNDNNGNVSLAQVRNDYIVFEGDGYTYQYDQWKATNKGADNIAMRDHVI 419


```

Db 313 IVQSEGRIYATKN--GVLGFCRTHQHLLVPDGGKIYQTRQ-----NPG-----VNFROPFTF 363
QY 420 EDGNG-----DRIYVFEA-----STGLNYQGEDQIYNWLNNGD-- 454
Db 364 RDQNPSPTEYMVFEGNSAFVREQQYVDAAKAGQNTTLATCTEEDLGYE-----KGDPK 419
QY 455 ----DAFNKSLFRILSNDIKSRATWANAAIGILKLNKDEKNPKVA--ELYSPLISAPM 508
Db 420 AETVEAVNRGGYYQLAN-----VGLAR-----AKNKAMTEWEYLPPLLSGNC 462
QY 509 VSDEIERPNVVKGNKYIYFAAATRLNRGSDNDAMNANAVG--DNVAMVGYVADSLTGS 566
Db 463 VNDOTERPOIYFDQGYLYFTISH-----RETYADGLOGPEGYGVFGDGLRSD 511
QY 567 YKPLN--DSGVVLTASVPANWRTA-----TYSYIYAVPVAKDDQVLVTSTM 610
Db 512 YKPLNQTGALGNPLNLFNFGKPYSPDNQSPYTFQSYSHYVMP--GGLVESEFIDSIG 569
QY 611 TNRNGVAGKGMSTWAPSLLOINPDNTTTLAKMTN 647
Db 570 GNKGDPNVRG--GSLSPYVKLNISGDTTSVDRTYCTN 604

RESULT 6
Q93FU9 ID Q93FU9 PRELIMINARY; PRT; 424 AA.
AC Q93FU9;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE LEVANSUCRASE.
GN LSCA.
OS Pseudomonas aurantiaca.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=86192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S-4380;
RA Koh I., Jang E.K., Kim J.S., Song K.S., Kim C.H., Rhee S.K.;
RT "Screening, molecular cloning and characterization of a novel
RL levanucrase and its expression in Escherichia coli.";
RL Submitted (Sep-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF306513; AAL09386.1;
SQ SEQUENCE 424 AA; 47009 MW; 9C1B4936A985162E CRC64;

```

```

Query Match 6.8%; Score 279.5; DB 2; Length 424;
Best Local Similarity 24.5%; Pred. No. 1.6e-05;
Matches 118; Conservative 76; Mismatches 174; Indels 113; Gaps 25;

QY 197 NKNMFSKAASGTMQYNDPQKI--ADTLIKQDGRYTVPFKASEIKNMPAATTKDAQWT 255
Db 4 NTEKFGAPHQPSLWTRADALKVHAD-----DPTTQPLVSA-----DFPVLSE----- 48
QY 256 IEPLDVNDSPVQDVRTGOVANNNGYQLVIMMG--IPNO-----NDNH 297
Db 49 --VFINDWPLKDL--DGNITSVDGWSVFTLTADRHPNDPEIDBNGYDITRDWNRH 104
QY 298 ----IYLLNKGDNLSLHWKNGVPIFGYNSTAVSQWEGSAY--LNSDNIQOLFTRVDT 352
Db 105 GRAKMYWFESRTGK----WKLGRVMAEGSVPTAREWAGTLPILLNEDGVDLYTAV-- 158
QY 353 SDNNTNHQIASATLYLTNNNGVSLAQVRNDYIVFEGDGYIYQTYDQ---WKATNKAGD 409
Db 159 ----TPCATIVKVRGVVTEHGVSMVGFEKVKPLFEADGKMYQEAQNPFW----- 206
QY 410 NIAMRDHVEDCN--GDRYLVFEASTGLNYQGEDQIYNWLNNGYDGAFAFNKSLFRILSN 468
Db 207 --GFRDPWPFDPNDGKLYLMEFEGNVAGE--RGSHKV-----GKAEIGDVPFGYEDVGN 256
QY 469 DDIKSRATWANAAIGILKLNKDEKNPKVAELYSPLISAPVSDIERPNVVKGNKYLYLF 528

```

```

Db 257 -----SREQTACVGI--AVARDEGDD--WEMPLPLLTAVGVNDQTERPHFVFDGKYLYLF 308
QY 529 AATRLNRGSDNDAMNANAVGDNV-----AMVGYVADSLTGSYKPLNDGSGVVLASVPAN 584
Db 309 T-----ISHTFTYADGVTPGDCYGVGFVADSLFGPYPLNGSLGLV--GNPSS 353
QY 585 WRTATSYIYAVPVAKDDQVLVTSTM-----TNRNGVAGKGMSTWAPSLLOINPDNTTT 640
Db 354 QPQFYSHCVMP-----NGLVTSFIDSPTDESQTQIR--IGGTEAPVEIKIKGQQTFFV 406
QY 641 V 641
Db 407 V 407

```

```

RESULT 7
Q9AEX9 ID Q9AEX9 PRELIMINARY; PRT; 415 AA.
AC Q9AEX9;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DE LEVANSUCRASE.
OS Rahnella aquatilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Rahnella.
OX NCBI_TaxID=34038;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim H., Cha J.;
RT "Cloning and characterization of levansucrase.";
RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY027657; AAK14794.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1_68.
SQ SEQUENCE 415 AA; 45958 MW; F09F5A1BF8690C5E CRC64;

```

```

Query Match 6.6%; Score 271.5; DB 2; Length 415;
Best Local Similarity 26.6%; Pred. No. 3.6e-05;
Matches 102; Conservative 60; Mismatches 132; Indels 89; Gaps 19;

QY 261 VMDSPVQDVRTGOVANNNGYQLVIMMGIPNQNNDHIYL-----LYNKYGDNE 309
Db 44 INDAMPLRSL--DGTVVSDGWSVFTLTQARN--NNNSETYLDADGNYDITSDNNRHGRK 101
QY 310 LSHWKN-----VGPFGYNSTAVSQWEGSAY--LNSDNIQOLFTRVDTSDNNTNHQ 360
Db 102 ICYWYSRTGKDWIFGGRVMAEGSVPTREWAGTLPILLNEDGVDLYTCV-----TPGA 155
QY 361 KIATATLYLTNDNNGVSLAQVRNDYIVFEGDGYIYQTYDQ---WKATNKAGDNIAMRD--A 416
Db 156 TIAKYRGKVLSEEGYVLGAFNEVKSLFSADGVYIYQTESQNTYWN-----FRDPS 205
QY 417 HVIEDGNGDRYLVFEASTGLNYQGEDQIYNWLNNGYDGAFAFNKSLFRILSNDIKSRAT 476
Db 206 PFIDPHDGLKLYMVEFEGNVAGE--RGSHVI-----GKQEMGTLPDHRDYG--AR 251
QY 477 WANAAGIL-----KLNKDEKNPKVAELYSPLISAPVSDIERPNVVKGNKYLYLFATRL 533
Db 252 YQAGCIGNAVAKDLSGDE-----WEILPPLVTAVGVNDQTERPHFVFDGKYLYFTISH-- 305
QY 534 NRGSDNDAMNANAVG-----DNVAMVGYVADSLTGSYKPLNDGSGVVLASVPANWRTAT 589
Db 306 -----KFTVAEGLTGPDGV--YGFSLDRLTGYPSPWNGSLGLV--GNPSPQFPQT 351
QY 590 YSYIYAVPVAKDDQVLVTSTM 612
Db 352 YSHCVMP-----NGLVTSFIDN 368

```

RESULT 8

[illegible]

```
QY 430 FEASTGLENYQGEDQIYNLNYGGDAFNKLSFLRILSNDDIKSRATWANAAGILKLNK 489
Db 467 ----TELYGYP-ERAYKSYVYGG-----YRLTWDNGL--VLYSNK 500
QY 490 DEKNPKVABL-----YSLISAPMVSEIERPNVVKLGNYLFAAATRLNRGS----- 537
Db 501 ADGNGKNGIIOQNDNFYKEDTAKGTMSQYDAKQIE-}-----TEENQDNPDLDI 550
QY 538 -----NDRAWNNANY--AVGNVAMVG----- 557
Db 551 DYHTAIDGGGVYGVYETIETDSADIDYHTAVDSEVHGVTESSESNPIDREE 610
QY 558 -----YVADSLTGSYKPLNDSGV-----LTASYPANWRTATYSY 593
Db 611 STHENSKHHADVVEEEDNPGGGQVTTESNLVEFDEESTKGIVTGAUSDHTTIEDTKEY 670
QY 594 AV-----PVAGKDDQVLVTSYMTNRNGVAGKGMDSWAPSLQINPDN 637
Db 671 TTESNLIELVDELPEHGAQAQPIEETENNHHISHSLGTENGHGNV--GVIEBIEENS 728
QY 638 TTTVLAKM-----TNOGDWIWDDSENLDN-----IGLDSAAALP----- 672
Db 729 HVDIKSELGYEGGQNGSQSFEEDTEE-DKPKYEGGNGNIVDIDFSDVPQIHGQNGQDSF 787
QY 673 ----GERDKP-----VWDLLIG--YGLKPHD----- 692
Db 788 EEDTEKDKPKYEHGGNIIDIDFSDVPQIHGFKHNEIEEDTNKDKPNYQFGGHSVDPE 847
QY 693 -----PATNDPET--PTTPTPETPTPKTPPEN 722
Db 848 EDTLPKVSQNGEQOTIETDTPPTPTPEVPSEPTPTPTPEVPSEPTPTPTPEVP 906
QY 723 PGTQPTPNTPNPEIPL-----TPETPKQETQT----- 751
Db 907 PSEPTP-TPPTPEVPSEPTPTPTPEVPSEPTPTPTPEVPSEPTPTPTPEVP 965
QY 752 -----NNRLPQTG--NNANKAMIGLGMGILLS 776
Db 966 PSKPEQGVVTPVIEINEKVKAVAPTKAQSKESELPETGGEESTNKGML---FGGLFS 1022
QY 777 MFLAEINKRRFN 789
Db 1023 ILGLALLRRNKN 1035

RESULT 10
Q53682 PRELIMINARY; PRT; 940 AA.
ID Q53682
AC Q53682;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN FNB8
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8325-4;
RX MEDLINE=92111475; PubMed=1837266;
RA Jonsson K., Signas C., Muller H.P., Lindberg M.;
RT "Two different genes encode fibronectin binding proteins in
RT Staphylococcus aureus. The complete nucleotide sequence and
RL characterization of the second gene.";
RL Eur. J. Biochem. 202;1041-1048(1991).
DR ENBL: X62992; CAA44726.1; -
DR InterPro: IPR004237; Fb_bind.
DR pfam: PF02986; Fb_bind; 2.
DR pfam: PF00746; Gram_pos_anchor; 1.
DR PROSITE: P500343; GRAM_POS_ANCHORING; UNKNOWN_1.
```

```
SQ SEQUENCE 940 AA; 103555 MW; E35FBBCA907AE345 CRC64;

Query Match 6.0%; Score 246; DB 2; Length 940;
Best Local Similarity 18.6%; Pred. No. 0.0013;
Matches 185; Conservative 135; Mismatches 334; Indels 340; Gaps 44;

QY 55 VTLSGSGQ-----VSAASDTTIRTSANANSASSAANTQNSOVASSAATSTSSAASLN 109
Db 25 IYVGGQKEAAASNNNTVEESGSSATESKASPTQTTNNVNT---IDETQSYSATST 81
QY 110 NPTDKAAQENTNTAKNDDTQKAAPANESSEAKNEPAVNVND--SSAANKDDQSS----- 162
Db 82 EQPSQSTQVTTTEAPK---TVQAPKVTESRV-DLPSEKVADEKTTGTQVDTIAQPSNVSEI 137
QY 163 ---KKNT--TAKLNKDAENVVKAGIDPNSLTD-----DOIK 194
Db 138 KPRMKRSTQVTAVAEKEVVEETKATGTDVTKNVEEGSEIVGHKQDVTNVVPHNAERTV 197
QY 195 ALNKMNFKAASGQTQNTYNDQKIADTLIKQDGRYTVPEFKASIKNMPAATTKDAQTN 254
Db 198 LKYKWKFGGKAG--DYEDF-TLSD-----NVETHGISLTKRVPEIKSTDGQV- 243
QY 255 TIEPLDVWDSHPQDVRTGQVANNV-----GYQLVIAMGIPNQN 294
Db 244 -----MATGEIIGERKVRITFKYVQEKKDLTAELSLSLNFIDPTTQK 288
QY 295 DNHIYLYNKYGDNELSHWKNVGPFGYNSTAVSOEKSGLAVLNSDNSIQLFYTRVDT-- 352
Db 289 NONVEV---KLGEITVSKIFNIQVLLG-----VRDNKGVT- - - - -NGRIDTLN 329
QY 353 -SDNNTNHOKTASATLYLTDNNGNVSLAQVNDYIVFEGDGYIYQTYDQWKATNKGADNI 411
Db 330 KVDGKFH-----FAYMKPNQSSLVTVTGQVTKGNKPCVNNPTVKVYK--HGSDDL 381
QY 412 A-----MRDAHVIEDGNDRLVFEASTGLE-NYQGEDQIYNML-----NYGGDDAFNIK 460
Db 382 AESVYAKLDDVSKEDYTDNMSLDFDNGYSLNFNLDOSKNYVIKYEYDYSNANLE 441
QY 461 SLRILS--NDDIKSRATWANAAGILKLNKDEKNPKVAELYSPLI--SAPMVSETERP 516
Db 442 FQHLFGYINYYSNLTWKN--GVAFYSNNAOGCKDLKEPIIEHSTPIELEFFSEP 498
QY 517 NVKLGKNKYLLFAATRLNRGSDDDAWNNY-----AVG----- 550
Db 499 PVKE-----HELTGTIEESNDSKPIDFEYHTAVEGAEGHAEGTIETEDSIHVDFEES 551
QY 551 -----DNVAMVGYVADSLTGSYKPLNDSGVV-----LTASYPANWRTATYSYA 594
Db 552 THENSKHHADVVEEEDNPGGGQVTTESNLVEFDEESTKGIVTGAUSDHTTIEDTKEYT 611
QY 595 V-----PVAGKDDQVLVTSYMTNRNGVAGKGMDSWAPSLQINPDNT 638
Db 612 TESNLIELVDELPEHGAQAQPIEETENNHHISHSLGTENGHGNV--GVIEBIEENS 669
QY 639 TTVLAKM-----TNOGDWIWDDSENLDN-----IGLDSAAALP----- 672
Db 670 VDIKSELGYEGGQNGSQSFEEDTEE-DKPKYEGGNGNIVDIDFSDVPQIHGQNGNQSFE 728
QY 673 --GERDKP-----VWDLLIG--YGLKPH----- 691
Db 729 EDEKDKPKYEGGNIIDIDFSDVPQIHGFKHNEIEEDTNKDKPNYQFGGHSVDPEE 788
QY 692 -----DPATNDPETPTTPTPETPTPKTPENPCTPQTPNTPN 733
Db 789 DTLPQVSGHNEGQOTIETDTPPTPTPEVPSEPTPTPTPEVPSEPTPTPTPEVPSEPTPT 846
QY 734 TPEIPLTP-----ETPKQPE-----TQNNRLPQ 757
Db 847 TPEVPTPEPKPIPPAKEEPKPSKPEQGVVTPVIEINEKVKAVVPTTKAQSKESELP 906
QY 758 TG--NNANKAMIGLGMGILLSMFLAEINKRRFN 789
```

Db 907 TGBESTNMGML---FGLFLSILGLALLRKNKN 937

RESULT 11

099051 PRELIMINARY; PRT; 1134 AA.

ID Q99051; AC Q99051; DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE IMMUNOGLOBULIN ALPHA FC RECEPTOR PRECURSOR.

GN BAC.

OS Streptococcus agalactiae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1311;

RN [1] SEQUENCE FROM N.A.

RP STRAIN=SB35;

RC MEDLINE=91257158; PubMed=2044657;

RX Heden L., Frithz E., Lindahl G.;

RA "Molecular characterization of an IgA receptor from group B streptococci: sequence of the gene, identification of a proline-rich region with unique structure and isolation of N-terminal fragments with IgA-binding capacity.";

RT Eur. J. Immunol. 20:2241-2247(1990).

RT -1- FUNCTION: PROTEIN BAC. IS A SURFACE PROTEIN WITH BOTH IGA-BINDING CAPABILITY AND A REGION CORRESPONDING TO THE BETA ANTIGEN.

CC EMBL: X58470; CAA41384.1; ;

DR PIR: A60234; A60234.

DR InterPro: IPR001899; Gram_pos_anchor.

DR InterPro: IPR003599; Ig.

DR Pfam: PF00746; Gram_pos_anchor; 1.

DR SMART: SM00409; IG; 1.

DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR REPEAT 826 915 30 TRIPEPTIDE REPEATS (MOTIF XPZ).

FT SIGNAL 1 38

FT CHAIN 39 1134 IMMUNOGLOBULIN ALPHA FC RECEPTOR

FT DOMAIN 1002 1007 CONSERVED IN GRAM-POSITIVE COCCI SURFACE

FT TRANSMEM 1111 1129 MEMBRANE ANCHOR REGION (BY SIMILARITY).

FT REPEAT 826 915 30 TRIPEPTIDE REPEATS (MOTIF XPZ).

SQ SEQUENCE 1134 AA; 127980 MW; FC10A200B3DB7CE1 CRC64;

Query Match

Best Local Similarity 5.8%; Score 237.5; DB 2; Length 1134;

Matches 189; Conservative 129; Mismatches 349; Indels 273; Gaps 39;

QY 6 KNAVVTLSAALVFCATTVNASADTNIENNDST-----VQVTTGNDIAVKSVTLGS 59

Db 42 KDSVATTEVAAPKY-----PSMAQTQDGNSSSELETTKMEIPTDIKKAPEPVKTA 96

QY 60 GOVSAASDTTIRTS-----NANSASSAANTONSNSOVASSAITSSTSAASL----- 108

Db 97 GETS-ATDTGKREKLOQKWNKLNKNDVNTILSHQKNEFTKIDETNDSOALLENQF 155

QY 109 NNT-----DSKAAQENNTAKNDTQ-----KAAAPANESSEAK- 141

Db 156 NETNRLHLHIKOEVEYKOKAKQKQ--KTLKQSDTKVDLSNIDKLNHQKQVKEKMAEQK 213

QY 142 --NEPAVNVNDSSAAKND-----QOSSKK-----NNTAKLNKDAENVV 178

Db 214 ITNED-----KDSMLAKKIEDIRKQAQQAOKKEDAEYKVREELGKLFSSTKAGLDQEIQHV 269

QY 179 KK-----AGID---PNSLTDQTKALNKMNFSSKAASGTMQTYNDF---QKIAD-- 221

Db 270 KKETSEENTQKVDEHYANSLQNLAKQSLLELDKATTNEQATQVK-NQFLENAQKLKEIQ 328

QY 222 TLIKQDGRYTVPFKA-----SEIKNMPAATTKAQTNTIPLDWDVS----- 264

Db 329 PLIKET--NVKLYKAMESLEQVEKELHNSLEANLDELVAKSEIYVREYEGKLNQSKNL 385

QY 265 -----WPQDVRTGOVANWNGYQ---LVIAMGI 290

Db 386 PELKOLEEEAHSLKQVVEDEFKFKTSEQVTPKKRYKRDLAANNENQCKTELTVSPENI 445

QY 291 PNONDHILLYLNKYGMELSHWKNVGFYGNSTAVSQEWSGSAVLNSDMSIQLFYTRV 350

Db 446 TVYEGEDVKFTVAKSDSKTT--LDFSLTLTKYNPSVSDRISTNKTNTDN-----HKIA 498

QY 351 DTSDDNTNHOKIASATLYLTDNNGNVSIAOVRNDYIVFEGDGYYYQYTDQWKATNKGADN 410

Db 499 EITIKNLKLNESQTVTLKAKDDSGNV--VEKFTTIV-----OKKEEKQVPKT 544

QY 411 IAMDRAHVIE-----DNGDRYLVFEASTGLENYQGEDQIYNMNYGGDDAFNIKS 461

Db 545 PEQKDSKTEEKVPOEPKSNQNLQELIKSAQOELEKLE-----KA 585

QY 462 LFRILSNDDIKS-----RATWAN-----AAIGILKLNKDEKNPKVAELYSPLISAP 507

Db 586 IKELMEQPEIPSNPEYGIQKSIWESQKEIOEAITSFKKIIGDSSSKYYTHYFNKYKSD 645

QY 508 MVSDEIERP-----NVKLGKNKYLLFAATRLNRGSDNDAMNANAVGDNVAMGVVAD 561

Db 646 FMNYQLHAQEMELTRKVVOYMNKYPDQNAEIKKIFESDMKTKEDNYGSLNDALKGYFEK 705

QY 562 SLTGSYKPL-----NDSGVLTASVPANWRATYSYAVPVAGKDDQ-----VLVTSYMT 611

Db 706 YELTPFNKIKQIVDDLDKVKVEQDQPA-----PIPENSEMDQAKKAKIAVSKYMS 755

QY 612 NRNGVAGKGMDSWAFSLLIQINPDNTTVLAKNTQGDWIWDDSSENLMDIGLDSAL 671

Db 756 -----KVLDDGVHQ-----HLQKKNHSKIV-----DLFKLELAIKQOTTFIDIINAKT 796

QY 672 PGERDKPVDWDLIG-----YGLK-----PHDPATPNDDPTPTPTPTPTPTPTPTPT 718

Db 797 EVIDNLVH-DAPSKMNAVAKFQKGLTNTPTPTPKIPQLPQAPDTPQAPDTPHVP 855

QY 719 TPENPGTPT 758

Db 856 SPRAPEAPRVPSPKTPEAPHVPSPKTPEAPKPIPEPKT 895

RESULT 12

Q91CJ7 PRELIMINARY; PRT; 1016 AA.

ID Q91CJ7; AC Q91CJ7; DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE DEXTRANSUCRASE.

GN DSRT.

OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OC Leuconostoc.

OX NCBI_TaxID=1245;

RN [1] SEQUENCE FROM N.A.

RC STRAIN=NRRL B-512F;

RA MEDLINE=20169623; PubMed=10705445;

RX Funane K., Mizuno K., Takahara H., Kobayashi M.;

RT "Gene encoding a dextranucrase-like protein in Leuconostoc mesenteroides NRRL B-512F.";

RL Biosci. Biotechnol. Biochem. 64:29-38(2000).

DR EMBL: AB020020; BAA90527.1; -

DR HSSP; P06278; IVJS.

DR InterPro; IPR003318; Glyco_hydro_70.

DR Pfam: PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1016 AA; 110344 MW; 8896EFDE13CCB47 CRC64;

```
Query Match          5.78; Score 236.5; DB 2; Length 1016;
Best Local Similarity 21.28; Pred. No. 0.0037;
Matches 162; Conservative 103; Mismatches 266; Indels 233; Gaps 37;

QY 1 MYKSGKNVAVTSLTAALVFCATVNASADTENN-----DSSTVQVTTGNDIA 51
DQ 1 MYKSGK---MLVIAGSVIIIGVTFIOQAQADVQNGVAVATVPQSNLDATTS- 53
QY 52 VKSVTLASGVQVSAASD-----TTTTSANASASSAANTONSNOVASSAATTSSTA 105
DQ 54 -KSITTDKKAATASTDDKATTADTSTDDKAAATAAT--STDDKATTATSTDDKAT 110
QY 106 ASLNTDTSKAA-----QENTNTAKNDTQKA---APANESSEAKNEPAVNVDSSA- 153
DQ 111 TAATSTDDKAAATTADTSTDDKATTATSTDDKATTATSTDDKATTATSTDDKAA 170
QY 154 ---AKNDQ-----OSSKNTTAKLKAENNVVKA-GIDPNSLTDDQIKALKNWFSK 203
DQ 171 TADTSTDDKATTATVGTSDNNNSTTASDKVSSSAQKSQIDNNSKTADTTAALEASSKNL 230
QY 204 AAKSGTQMTVNDFOKIADTLIKODGRYTPFPKASEIKNMPAATTKDAQTNTIEPLD 263
DQ 231 KTIDGRTYYDD-----DQVK-----KNF--ATVIDGKVLVF----- 261
QY 264 SWPQVQRTGOVANWNGYQLVIAMGIPNQNDNHLYLLNYKGDNELSHWKNVGFIFYN 323
DQ 262 -----DKETGALADTNDYQFLEGL-----TSENNTYTHNASVGTSTSYTNVD--GY- 307
QY 324 STAVS-----QEHSGSAVLNSNSIQLFYTRVDTSDNNNTNHQIKASATLYLTDN 372
DQ 308 LTADSHYRPFKILVNGQNWESSKDDLRPLMTWMPDKATQVNYLNAMKYLDAETET- 365
QY 373 NGNVSLAQVRNDYIVFEGDGYIYQYQDWKATNKGADNIAMDARHVEDGNGDRYL- 428
DQ 366 -----VYTSDDSDQALNKAQAQNIQVRIEKSISQEGTQWMLKDDI 404
QY 429 -----VFESTGLENYQGEDQYINWNLNGYGGDDAFNKLFLRLSNDIDKSRATW 477
DQ 405 SKFVDSQSNWNIASEKSTHLOGGALLY-----VNSDKTPDANSYRL----- 448
QY 478 ANAATGILKLNKDEKNPKVAELYSPLISAP-----MVSDETERPNV-----KLGKYY 526
DQ 449 -----LNRTPNQTGTPLYT--TDPTQGYDFLLANDVNSPNVVOAQLANWY 496
QY 527 LFAATRLNRGS--NDDAWMANY-----AVGDNV--AMGYVADSLTGSYKPLNDSGVV 576
DQ 497 L-----LNFGSITNDA--DANFDSIRVDV-DNVADLLQIAADYFKAAY-GVDKSDAI 547
QY 577 LT--ASVPANRRTATSYIYAVPVAGKDDQVL-----VTSYWTNRNGV 616
DQ 548 SNQHVSILEDWSDNDAY-----VKDNGDNQLSDMKLRLSLKSLYMPAVDQYGNKRSGL 603
QY 617 -----AGKGMSTWAPSF-LQINPDNTTTLVAKMTNQ 648
DQ 604 EPFLTNSLVDRNTDNTDNTAQPNIYFVRAHDSVQTVIAELIKQ 647
```

RESULT 13
Q99RD3 ID Q99RD3 PRELIMINARY; PRT: 961 AA.
AC Q99RD3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FNBB PROTEIN (FIBRONECTIN-BINDING PROTEIN HOMOLOG).
GN FNBB OR SA2290 OR SA22502.
OS Staphylococcus aureus (strain N315), and
GS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879, 158878;

```
[1]
SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.:
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003137; BAB43593.1; --
DR EMBL; AP003365; BAB58664.1; --
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 961 AA; 106010 MW; 364940F884EA4101 CRC64;
```

Query Match 5.68; Score 231.5; DB 16; Length 961;
Best Local Similarity 19.98; Pred. No. 0.0059;
Matches 199; Conservative 121; Mismatches 349; Indels 331; Gaps 52;

```
QY 55 VTGSGQ-----VSAASDPTTIRTSANASASSAANTONSNOVASSAATTSSTSSAASLN 109
DQ 25 IVVGMGQKEKAAASEQNNTTVEESGSSATESKASEQTQTTNNVNT---IDETQSYSATST 81
QY 110 NTDSKAAQENTNTAKNDTQKAAPANESSEAKNE--PAVNVD-----SSAA 154
DQ 82 EQPSKSTQVTEEA---PTTVQAPKVETEMKSOEDLPSEKADVADKETTQVQDIAQPSNV 138
QY 155 KNDQDQSSKNTTAKLKAENNVVKKAGIDPNS---LTDDQIKALN----- 197
DQ 139 EIKPRMKRSADVATVSEKEVAEEAKATGDTVTKNKEVTESSLEGHNKNSINVYVPHNAORV 198
QY 198 ---KNFESKAQSGQMTYNDFOKIADTLIKODGRYT--VPFFKAS-EIKNM----- 243
DQ 199 TLKYKFKFEGEIKAG---DYDFD-TLSDN-VETHGISTLRKVPKESKTEDEKVMANGQVI 253
QY 244 -----PAATTKD-----AQTNTIEPLDWDSD 264
DQ 254 NERTIRYTFDYINNKKDLTAELNLNLFIDPTVTYKQSQKVEVTLGQNKVSKFEIDIKYL 313
QY 265 WPDQVQRTGOVANWNGYQLVIAMGIPNQNDNHLYLLNYKGDNELSHWKNVGPY-FGYN 323
DQ 314 DGKVD-RMGVTVNGR-----IDTLNKEEGKFSHFAYKPKNNQSLTSVTYVGTQVTSYK 365
QY 324 STAVSQEWSSGSAVLNSNSIQLFYTRVDTSDNNNTNHQIKASATLYLTDNNG-NVSLAQVR 382
DQ 366 QSANNPTVKYKHIGSDLAESVYAKLD--DTSKFEDVTEKVNLSYTSNGGYTLNLGDLD 423
QY 383 N-DYIV-FEGD-----GY-----YYQTYDQWKATNKA---DNIAM 413
DQ 424 NSKDYIKYEGEYDONAKDLNFRTHLSGYHKYYPYPPYPPYVQLTWNNGVAFYSNNAK 483
QY 414 RDAH-----VLEDGNGDRYLVFESTGLENYQ- 441
DQ 484 GDGKDKPNPDIKSEPIDDIKSEPPVEKHELTGTTEESNDSKPIDFYHTAVEAGEGH 543
QY 442 -----EQIY-----NWLNYGGDDA-----FNISLFR 464
DQ 544 AEGIIETEEDSIHVDFEESTHENSKKHADYVEEDTNPQGGQVTTESNLVEDEESTKG 603
QY 465 ILSN-----DDIKSRATWANAAIGILKLNKDEKNPKVAELYSPLISAPVSDIERP 516
DQ 465 ILSN-----DDIKSRATWANAAIGILKLNKDEKNPKVAELYSPLISAPVSDIERP 516
```

```

Db 604 IVTGAVSHTTVEDTKEYTESN-----LIEL-VDELPEEHQAQGP-----EETEN 651
QY 517 N-----VYKLGKYYLFAATRLNRSGNDNANAVGDNVAMVGVVADSLTGSYK 568
Db 652 NHHISHSGLTENGHN--GVIDEIENSHVDIKSELGEGQNGSQSFEED--TEEDK 708
QY 569 PLNDGSGVLTASVPANWRTATYSYAVP-VAGKDDOVLVTSYWNRRNGVAGKGMDSW-A 626
Db 709 PKYEQG-----GNVIDIDSDVPQJHGQ-----NGNSFEEDTEEDK 746
QY 627 PSF-----LLQINPDNTTTLAKMNGQDWIDSS-----NLDWIG-----DLDLAAAP-- 672
Db 747 PKYEQGNIIDIDSDVPQJHG-FNKHNEIIEEDTKDKPNYQFGHNSVFEEDLPKV 805
QY 673 ---GERDKPVDWLDIGYGLKPHDPATPNDEPTTPTETPTNTP--KTPKTPENPGTPQ 727
Db 806 SGQNEGOQTIEEDT--PTPTPTPEVSEPTP--TPPTPEVSEPTP--TPPTPEVSEPTP 862
QY 728 TPNTPTPEIPLP-----ETPKOP-----ETQT 751
Db 863 TP-VPTPEVSEPTPPEVPAKEPKPKPVQEGKVVTPVIEINEKYKAVAPTQKOSK 921
QY 752 NNRLPOTG--NNANKAMIGLGMGTLLSMEGLAEINKRRN 789
Db 922 KSELPTGGEESTNKML---EGGLFSLGLVLLRRNKN 958

RESULT 14
Q9KFW0
ID Q9KFW0 PRELIMINARY; PRT; 1661 AA.
AC Q9KFW0;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE BH0361 PROTEIN.
GN BH0361
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN
RP
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001508; BAB04080.1;
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1661 AA; 183657 MW; 837658930D2E60C4 CRC64;

Query Match
Best Local Similarity 5.6%; Score 230; DB 16; Length 1661;
Matches 181; Conservative 110; Mismatches 288; Indels 310; Gaps 41;

QY 92 VASSAAITSTSSAASLNNSTSKAAQENTAKNDTOKAAPANESSAKNEPAAVNVNDS 151
Db 885 IDSAYWTFKT-----SLEDTLKSNYPNATLKSD-----NAEDFNVDAS 925
QY 152 SAAKNDQOQSSKNWTAKLNKDAENVYKAGIDPNSLTDDQIKALNKNFSKRAKS--- 208
Db 926 VSVRRGGE-----FVTKGAQNGRLDWSL-----NINFSQSSISNAKI 964
QY 209 -----TQMTVNDFOKIADTLIKODGYTPVFFKASEI---KNMPAATTKDAQTNTIEPL 259
Db 965 VDPSVNVILLIEDTFRLYATNVQPNGTWT-----KGDELERDKOYTLDIRDTGEGNQHFEL 1020

```

```

QY 260 DVWDSNPVODVRTQGVANNGYQLVTAMGIPNQNNDNHIYLYNKYKGNLSHMKNVGP 319
Db 1021 QF-----TEIDITGYILEYQSF--INARNG--DVVTNNVKGLEGDLTTEEYVE----- 1064
QY 320 FGYNSTAVSQEWSGSAVLNSDNSIQLFYTRVTRDSDNNTNHQKIASATYLYLDNNGNVSLA 379
Db 1065 ---SSGAVSVRLSGAGTSGTGSLEVTKVDADTG-----EVLOGATFTLYDSEGEFAIR 1117
QY 380 QVR-----NDYIVFEG-----DGYYYOTYDQWKATNKGADNIAMRAH--VIE 420
Db 1118 TLETGEDSKATFVNLLYGYLLKEDSAPEGLVGINDTQRTV---IDTV---LHEVTVE 1170
QY 421 DONGDRYLV-----FEASTGLENYQGEDQIYNKLNNGGDDAFNKSIFRILSNDD 470
Db 1171 NEKSDINRVSAVAGVQLQKVDDETG-ESLOG-----ALFALQOKVD 1210
QY 471 IKSRTAWANAAGITLKLKNDK-----NPKVAELYS-----LISAPMV-SDEIE 514
Db 1211 -----DEFVTTAEMETDEEGIVFAGSLEPGDYQFVELNAPVGYKLDPTVWTFVEED 1262
QY 515 RPNVVKLGKNYLFAATRLNRSGNDNANAVG-----DNVAMVGVVADSLT 564
Db 1263 RTETIELQKENHLIPGSQLVKVDADA---ANTLEGAEFTLLDGEVNVQEGSLTTD--- 1316
QY 565 GSYKPLNDSGVLTASVPANWR-----TATYSYAVPV----- 597
Db 1317 -----ENGQVVVTDLKPGEQYFVETKAPAGYELEATPIGFTIERNOQEVATVAVENHLI 1370
QY 598 -----AGKDD-----QVLVTS-----YMTNR 613
Db 1371 PGSQLVKVDADAATLEGAEFTLLDGEVNVQEGSLTTDATGQVVVTDLKPGEQYFVET 1430
QY 614 NGVAGKGMDSWAPSLLOINPDNTTV-----LAKMNOQGDWIDSSSE--N 659
Db 1431 KAPAGYELDAT--PIGFTIERNOQEVATVAVENHLIPGSQLVKVDREDSAVTLEGAEFSL 1489
QY 660 LDMIGD-LDSALPGERD-----KPVWDWLI-----GV----- 686
Db 1490 LDEEGNVLRGIRTGENGQMLVIDLKPQDYQFIEFKATGYELDDTPIAFIEKGTDEVL 1549
QY 687 -----GLKPHDPATPNDEPTTPTETPTNTPKTPKTPENPGTPQTPNTPTPEIPLT 740
Db 1550 TLVFNKLTPEVSEPTENPENPENPEIDLETPESENPETPEPKHSESEVPSNS 1609
QY 741 PETPKQPTQTN--NRLPOTGNANKAMIGLGMGTLLSMEGLAEINKRR 787
Db 1610 DKNPEKSSDNNRIDRLPQTGEEFTLLILGL--LLMTAGGVLLKARR 1656

RESULT 15
Q9LOA7
ID Q9LOA7 PRELIMINARY; PRT; 138 AA.
AC Q9LOA7;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE F4N2.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Dunn P., Khan S., Kim C.,
RA Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A., Hansen N.F.,
RA Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S., Luros S.,
RA Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G., Davis R.W.,
RA Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F4N2 from chromosome
RT I.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

```

	Query Match	5.4%	Score 224;	DB 10;	Length 138;
	Best Local Similarity	60.9%	pred. No. 0.0015;		
	Matches 42;	Conservative	0;	Mismatches 27;	Indels 0;
	Gaps	0;			
QY	690	PHDPATPNDETPTTPTETPTNTPKTKTPENPGPQPTNTPTNTPTPLTPTPKQPET	749		
Db	33	PGIPATPTATPATPTATPTNTPTTTTPTTPTPATPATPATNTPLTPTTPTTTT	92		
QY	750	QTNNRLPQT	758		
Db	93	PTTTLTPTT	101		

Search completed: September 26, 2002, 18:27:55
Job time: 254 sec

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	2016.5	48.8	797	1	SACB_STRMU	P11701 streptococc
2	1739.5	42.1	969	1	SACB_STRSL	O55242 streptococc
3	744	18.0	473	1	SACB_BACSU	P05655 bacillus su
4	741.5	18.0	472	1	SACB_BAGAM	P21130 bacillus am
5	721	17.5	473	1	SACB_BAGST	P94468 bacillus st
6	304	7.4	423	1	SACB_ZYMO	O60114 zymomonas m
7	301.5	7.3	413	1	INVB_ZYMO	O60115 zymomonas m
8	281	6.8	415	1	SACB_ERHAM	O46654 erwinia am
9	275.5	6.7	415	1	SACB_RAHQA	O54435 rahnella aq
10	274	6.6	431	1	SACB_PRESH	O68609 pseudomonas
11	266	6.4	584	1	SACB_ACEDI	O452408 pseudomonas
12	257	6.2	415	1	SACB_PESGG	O53998 acetobacter
13	255	6.2	1018	1	FNBA_STRAU	P14738 staphylococ
14	241	5.8	1164	1	BAG_STRAG	P27951 streptococc
15	229.5	5.6	1500	1	SSP5_STRGN	P16952 streptococc
16	227	5.5	1337	1	DEXT_STRDO	P39653 streptococc
17	220	5.3	826	1	SSP2_PLAYO	Q01443 plasmodium
18	213	15.2	1565	1	PAC_STRMU	P11657 streptococc
19	211.5	5.1	850	1	DEXT_STRMU	Q54443 streptococc
20	208.5	5.0	1256	1	MRP_STRSU	P32653 streptococc
21	205.5	5.0	1637	1	MRSP_STRAU	P80544 staphylococ
22	199	4.8	1592	1	GTF2_STRDO	P27470 streptococc
23	195	4.7	1758	1	YIR7_YEAST	P40434 saccharomyc
24	195	4.6	1561	1	YJW5_YEAST	P40889 saccharomyc
25	191.5	4.7	1561	1	SPAP_STRMU	P23504 streptococc
26	190	4.6	1597	1	GTF1_STRDO	P11001 streptococc
27	185.5	4.5	1528	1	SPAA_STRDO	P21979 streptococc
28	185	4.5	1569	1	YFUA_ECOLI	P52143 escherichia
29	184	4.5	1462	1	GTFD_STRMU	P49331 streptococc
30	183.5	4.4	881	1	YJH8_YEAST	P47033 saccharomyc
31	183.5	4.4	1694	1	IGAO_HAETN	P44969 haemophilus
32	182.5	4.4	1296	1	ASAI_ENTFA	P17953 enterococcu
33	182.5	4.4	1396	1	VLTFF_BP15	P13390 bacteriopho

RESULT	2
SACB_STRSL	
ID	SACB_STRSL STANDARD; PRT; 969 AA.
AC	Q55242;
DT	15-DEC-1998 (Rel. 37, Created)
DT	15-DEC-1998 (Rel. 37, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
DE	D transferase) (Sucrose 6-fructosyl transferase).
GN	PFF.
OS	Streptococcus salivarius.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
CC	Streptococcus.
OX	NCBI_TaxID=1304;
RN	[1]
SEQUENCE FROM N.A.	
RP	STRAIN-ATCC 25975;
RC	MEDJUNE-93322332; PubMed-8331080;
RX	Rathsam C., Giffard P.M., Jacques N.A. ;
RT	"The cell-bound fructosyltransferase of Streptococcus salivarius: the
RA	carboxyl terminus specifies attachment in a Streptococcus gordonii
RE	model system".
RZ	

2

DR	PIR:	A25040;	A25040.
DR	PIR:	S07309;	S07309.
DR	Subtilist:	BGI0388;	sacB.
DR	InterPro:	IPR003469;	Glyco_hydro_68.
DR	Pfam:	PF02435;	Glyco_hydro_68; 1
KW	Transferase:	Glycosyltransferase;	Signal; Complete proteome.
FT	TRANSFAL	1	29
FT	CHAIN	30	473
FT	CONFLICT	12	12
FT	SEQUENCE	473 AA;	53971 MW; 3FBF2F571B41D5B0 CRC64;
SQ			LEVANSUCRASE. V -> I (IN REF. 3).

Query Match 18.0%; Score 744; DB 1; Length 473;
Best Local Similarity 36.8%;
Matches 183; Conservative 97.4%;
Pred. No. 2.4e-30;

```

QY      199 MNFSKAAKSGTQWTY-----NDFOKIALA
       || | | | :|
Db     1 MNIKFAKQATVLTFTALLAGGATAQAFKETNOKPYKETYGISHTRHMDLQIPIB

```

Qy	226	QGRYVTPFFKASEIKNPAATTKDAQNTIEPLDWDWSHPVDVRTGOVANNNG
		::: :::
Db	59	KNEKYQVPEDSDSTIKNISSA-----KGLDWDWSPLQNA-DGTVANYHGY
		: :::

```

      .286 AMMGIP-NQNDNHHYLLYNKYGDNELSHWKNVGPI-----FGYNSTAV---SQEW
            | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db         108 ALAGDPKKNADDTSYMEYOKVCEPMSICGCTC-----

```

336 VLNSDINSIQLFYTRVDTSDNNTNHQKIASATLYLTLDNNGNVSLAQVRNDYIVFEGD
 QY

168 TETSDGKIRLEYT--DFSGKHGKQTLTTAQVNVASDSSLNNGVINGVEDYKSFIDGD

350 QIUDW--KATNKGADNIAMRDAHVIEDNGDRYLFVEASTGLEN-YQGEDQIYNW
| | : | : || | : || | : ||||| : || : |||| : : |
226 QNVQQFIDEGNYSSTDNHTLRDPHYVED-KCHUKYVERAUMCTHON-
Db

453 GDDAFNIKSLFRILSNDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPMM

db 285 KSTSFRRQESQLLQSDK-KRTAELANGALGMIELNDDYTLKKVMK---PLIASNT
 513 IERPNNVWIKICNVKVIETATMGANSGE

341 IERANVFKMNGKWYLFDTD---SRGSK---MTIDGITSNDIYMIQVYVNSITCQVYK

573 SGVLTASVPANWRTATSYAYAVPVAGKDDQVLVTSYMTNRNGVAGKGMDSWAPSE

394 TGLVLKMDLDPNDVTFTYSHFAVPQA-KGNVVVITSYMTNRGFYADK--QSFAPSEF
633 INPDNTTTLVLAKMTNOC 640

b 451 IKGKKTSVVKDSILEQG 467

RESULT 4

SACB_BACAM	
D	SACB_BACAM
C	P21130;
	STANDARD;
	PRT; 472 AA.

01-FEB-1991 (Rel. 17, Created)
01-FEB-1991 (Rel. 17, Last sequence update)
15-DEC-1998 (Rel. 27, Last sequence update)

Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase) (Sucrose 6-fructosyl transferase).
CNCB

Bacillus amyloliquefaciens.
Bacteria; Firmicutes; Bacillus/Clostridium

Bacillus/Staphylococcus group; Bacillus.
NCBI_TaxID=1390;
[1]

SEQUENCE FROM N.A.
STRAIN=ATCC 23844;
MEDLINE=01002500

Tang L.B., Lenstra R., Borchert T.V., Vasantha N.;
"Isolation and characterization of levanucrase-coding

----- of transucrase-encoding gene

```

Query Match      17.5%   Score 721, DB 1;   Length 473;
Best Local Similarity 36.3%   Pred. NO. 3.3e-29;
Matches 181; Conservative
199 MFNSKAASGQTMYY-----NDFOKIADTLIK 225
Db 1 MNIKFAKQATVLTFTALLAGGATQAFAKETNQKPYKETYGISHITRDMLOIQEQ--Q 58
QY 226 QDGRYTVPPFKASEIKNMPAAATTKDAQNTIEPLDWDSPVQDVRTGOVANNWGQYQVLI 285
Db 59 KNEKYQVPEFDSSTIKNISA-----KGLDWDMSWPLONA-DCTVANYGHYHVF 107
QY 286 AMMGIP-NQNDNHIIYLLNKKYGDNELSHKNVGPI-----FGYNSTAV-----SQEWSGSA 335
Db 108 ALAGDPKNADDTSIYWFYQKVGTSDISWSTKPGRYFKDSKDFDANDSLTKDQTOEWSSGA 167
QY 336 VLNSDNSIOIIFYTRVDTSONNTHOKIASATILYLDNNGNYSLAQVRNDYIVFEGDYYY 395
Db 168 TFTSDGKIRLFYF--DFSGKHYGKQTLTAQVNVWSASDSSINGVEDYKSIFFGDSKTY 225
QY 396 QTYDQW--KATNKGADNIAMRAHVIEDGNGDRILYVFEASTGLEN-YQGEDQIYNNMLNYG 452
Db 226 QNYQQQFIDBNGSSGNDHTLRDPHYVED-KGHKVLVFPEATGTEDGYQGESLFNKAYYG 284
QY 453 GDDAFNIKSLFRILNSDDIKSR-ATWANAAGIILKNRDEKNPKVAELYSPLIASPWSD 511
Db 285 KSTSPFQESQKLQSD--KNRTAELANGALGMTELNDYVTLKKVKM---PLIASNTVTD 339
QY 512 ETERPNNVKLGNNKYLYFAATRLNRGSDNDAMWNNANYAGDVNAWGVYVADSUTGSKYKPLN 571
Db 340 ETERPNNVKMGKWL-----MTDGTISNDIYMLGVNSLSTGPKYKPLN 392

```

QY 572 DSGVLTASYPANRTATSYAVPVAGKDDQVLVTSTNTRNGVAGKGMDSWAPSLL 631
 Db 393 KTLGLVKMLDNDVFTTYSFAVPOA-TGNVVTSTNTRNGFYADK--QSTFAPSLL 449
 QY 632 QINPDNTTTLAKMTNOG 649
 Db 450 NIQGRKTSVVRASILDQ 467

RESULT 6
 SABC_ZYMMO
 ID SABC_ZYMMO STANDARD; PRT; 423 AA.
 AC Q60114; Q60116; Q06487;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Levanucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
 GN SABC OR LEVU OR SUCE2.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 10988 / ZM1;
 RX MEDLINE=93305726; PubMed=8318541;
 RA Song K.B., Joo H.K., Rhee S.-K.;
 RT "Nucleotide sequence of levansucrase gene (levU) of Zymomonas mobilis
 RT ZM1 (ATCC10988).";
 RL Biochim. Biophys. Acta 1173:320-324(1993).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=IFO 13756;
 RX MEDLINE=95218269; PubMed=7766026;
 RA Kyono K., Yanase H., Tonomura K., Kawasaki H., Sakai T.;
 RT "Cloning and characterization of Zymomonas mobilis genes encoding
 RT extracellular levansucrase and invertase.";
 RL Biosci. Biotechnol. Biochem. 59:289-293(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL B806;
 RX Gunasekaran P., Mukundan G., Kannan R., Velmurugan S.,
 RA Ali-Abdelkader N., Alvarez E., Baratt J.;
 RT Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 31821 / ZM4 / CP4;
 RX Ahn J.Y., Kang H.S.;
 RT "Sequence analysis of 4486 fosmid clone of Zymomonas mobilis ZM4.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
 CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
 CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- PTM: DOES NOT SEEM TO BE N-TERMINALLY PROCESSED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF081588; AAA27695.1;
 CC EMBL; D17524; BAA0475.1;
 CC EMBL; L33402; AAA27702.1;
 CC EMBL; AF313764; AAC29870.1;
 CC InterPro: IPR003469; Glyco_hydro_68.
 CC Pfam: PF02435; Glyco_hydro_68; 1.
 CC Transferase; Glycosyltransferase.

FT CONFLICT 39 39
 FT CONFLICT 200 200
 FT CONFLICT 217 217
 FT CONFLICT 220 220
 FT CONFLICT 244 244
 FT CONFLICT 379 379
 SQ SEQUENCE 423 AA; 46762 MW; 19A4691DA3EB9FE3 CRC64;

Query Match
 Best Local Similarity 7.4%; Score 304; DB 1; Length 423;
 Matches 120; Conservative 74; Mismatches 155; Indels 102; Gaps 23;

QY 250 DAQNTNIEPLD-----VWDSHPVQDVRTGQVANNNGYQVLTAM-----GIN 292
 Db 24 DDPATMTIDYDFPVMTDKYVWDTWPLRDI-NGOVVSFGQSVIFALVADRKYGVHIN 82
 QY 293 QNDN-HIYLLYKNGDNLHSHKNGVGFYGNSTAVSQWSSAVL--NSDNSIQLFYTR 349
 Db 83 RNDGARIGYFSRGSN-----WIFGGHLLKDGANPRSWSGCTINAPATANSVEVFTS 138
 QY 350 VQTSDDNNTNHQKI-ASATLYLTNNNGVSLAQRNDYIVFEGDGYGYQTYDQWKATNKA 408
 Db 139 V-----NDTPSESVPAOCKGYIADKSWFDFGDKVTDLFOADGLYADY-----AENFW 190
 QY 409 DNTAMRDAHV-IDGNGDRYLVEASTGLENYQGEDQIYNWLNNGDDAFNIKSLFRILS 467
 Db 191 D---FRDPHFVFNEDGKTYALFEGVAMER-----GTAVAGEEIGPVPP 233
 QY 468 NDDIKSRATWANAAGILK-LNKDEKNPKVAELYSPLISAPMVSDEIRPNVWKLGNKY 526
 Db 234 KTETPDGARYCAAGIAQALNEARTEWR---LLPPLVAFGVNDQTERPHVFPQGLTY 290
 QY 527 LFAAFLNRSNDAMNANTAVG---DNVAMVGYVADS-LTGSYKPLNDGCVLTASV 581
 Db 291 LFTISH-----HSTYADGLSGPDGV--YGFVSENGIFGYPFPLNGSLGLV--GN 335
 QY 582 PANWRTATSYAVPVAGKDDQVLVTSTN-----RNGVAGKGMDSWAPSFLQ 632
 Db 336 PSSQPYQAYSHVM-----TNGLVTSFDTIPSSDPNNVRYG-----GTLAPTIKLE 382
 QY 633 INPDNTTTLAK-----MTNQGDWTDWSSSEN 659
 Db 383 LVGHRSFTEVKGYGYIPPOLEWLADESSN 413

RESULT 7

INVB_ZYMMO
 ID INVB_ZYMMO STANDARD; PRT; 413 AA.
 AC Q60115; Q60117; Q60125;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Extracellular sucrose (EC 3.2.1.26) (Beta-fructofuranosidase)
 GN SACC OR INVB OR SUCE3.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 10988 / ZM1;
 RX MEDLINE=94368848; PubMed=8086457;
 RA Song K.B., Lee S.K., Joo H.K., Rhee S.-K.;
 RT "Nucleotide and derived amino acid sequences of an extracellular
 RT sucrose gene (invB) of Zymomonas mobilis ZM1 (ATCC10988).";
 RL Biochim. Biophys. Acta 1219:163-166(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IFO 13756;
 RX MEDLINE=95218269; PubMed=7766026;
 RA Kyono K., Yanase H., Tonomura K., Kawasaki H., Sakai T.;

ISHKYTFADNLTPDGVYGFVSDKLTGPTYPMNSSGLVL--G 342

LDV L V I S I M T N --- R N G -- V A G K G M D S T W A P S E L L Q I N P D N T 638

298 Y Y L F T ----- I S H K Y T F A D N L T G P D G V Y G F V S D K L T G P Y T P M N S S G L

```

QY 501 VPANRRTATSYAYVPVAGKDDQVLVTYMTNRNGVAGKMD-----STWAPSFLLQINPD 636
Db 343 NPSSQPFQFTSYHYMP-----NGLVTSFI---DSVPWKGKDYRIGGTTEAPTVKILLKGD 393
QY 637 NTTTTV 641
Db 394 RSFTV 398

RESULT 9
ID SABC_RAHQAQ
AC 054435; STANDARD; PRT; 415 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
DE (Sucrose 6-fructosyl transferase).
GN SACS.
OS Rahnella aquatilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Rahnella.
OX NCBI_TaxID=34038;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33071;
RX MEDLINE=99127094; PubMed=9928133;
RA Song K.B., Seo J.W., Kim M.G., Rhee S.K.;
RT "Levansucrase of Rahnella aquatilis ATCC33071. Gene cloning,
RL expression, and levan formation.";
RL Ann. N.Y. Acad. Sci. 864:506-511(1998).
CC 1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC 1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commer-
CC cial entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U91484; AAC36458.1;
CC InterPro: IPR003469; Glyco_hydro.68.
CC Pfam: PF02435; Glyco_hydro.68.1.
CC Transferase; Glycosyltransferase.
CC SEQUENCE 415 AA; 45939 MW; E028828813D13A74 CRG64;

```

Query Match	6.7%	Score	275.5;	DB 1;	Length	415;
Best Local Similarity	27.2%	Pred. No.	3.9e-07;			
Matches	104;	Conservative	56;	Mismatches	132;	Indels
89;	Gaps	19;				
Qy	261	VWDSWPEQDVRFQVANNNGYQLVIAMGPIPNQNDHIYL-----LYNKYGDNE	309			
Db	44	IWDTPMLRSL-DGTVVSVDGMSVIFTLTAQRN-NNSEYLDIAEGNYDTSQWNHRGRAR	101			
Qy	310	LSHWKN-----VGPIFGVNSTAVSOEWSSGSAV-LNSDSIQLFYTRVDTSDNNTHQ	360			
Db	102	ICYWTSRTGKDWIFGGRVMAEGVSPSREWAGTPILLNEDGIDIIYYTCV-----TPGA	155			
Qy	361	KIASATLYLTDDNNGNSVLAQVRNDYIVFEGDGYYYQTYDQ---WKATNKGADNIAMRD-A	416			
Db	156	TIAKVRGKVLTSSEGVTLACFNEVKLSFGADGYYYQTESQNPYN-----FDPDS	205			
Qy	417	HVIEDGCDRLYFPEASTGLENYQGEDQIYNWLNLYGGDDAFENIKSLFRILSNDDIKSRAT	476			
Db	206	PFIDPHDGKLYMYFEGNVAGE--RGSHVI-----GKQMGTLPPGHRDVGN-----AR	251			
Qy	477	WANAAGIL---KLNKDEKNPKVAELYSPLISAPWSDETERPNVVKLGKNYYFAATRL	533			

Db	252	YQACIGMAVAKDL	SGDE-----WEILLP	LYATGVNDQTERPHFVFQDKYYLFT	ISH- 305
QY	534	NRGSNDDAWMNANT	AVG-----DNNAMVGYVADSLT	SGYKPLNDSGVVLTA	SPVANNWRTAT 589
Db	306	-----KFTVADGLTG	DGV--YGLFSLDNL	TGTPSPMNGSLVL--GNPPSQ	PQFOT 351
QY	590	YSYAYPVPVAGKDDQ	VLVTSYNTN 612		
Db	352	YSHCVMP-----NGLVTS	FDIN 368		
RESULT 10					
SACB_PSESH					
ID	SACB_PSESH	STANDARD;		PRT;	431 AA.
AC	O68609;				
DT	15-DEC-1998	(Rel. 37,	Created)		
DT	15-DEC-1998	(Rel. 37,	Last sequence update)		
DT	15-DEC-1998	(Rel. 37,	Last annotation update)		
DE	Levansucrase (EC 2.4.1.10)	(Beta-D-fructofuranosyl transferase)			
GN	USC	(Sucrose 6-fructosyl transferase)			

```

RESULT 10
SACB_PSSH
ID      SACB_PSSH      STANDARD;      PRT;      431 AA.
AC      068609;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      levanucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
DE      (Sucrose 6-fructosyl transferase).
GN      LSC.
OS      Pseudomonas syringae (pv. phaseolicola).
OC      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX      Pseudomonas.
NX      NCBI_TaxID=319;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=NCPPB 1321;
RX      MEDLINE=98394981; PubMed=9726857;
RA      Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
RT      "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT      levanucrase genes from the plant pathogens Pseudomonas syringae pv.
RT      glycinea and P. syringae pv. phaseolicola";
RL      Appl. Environ. Microbiol. 64:3180-3187(1998).
CC      -!- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
CC      GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC      ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC      -!- SUBCELLULAR LOCATION: Secreted.
CC      -!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----

```

DR EMBL; AF052289; AAC36063.1; -
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 431 AA; 47603 MW; 795FEA246D40C40F C5C64.

```

Query Match      6.6%; Score 274; DB 1; Length 431;
Best Local Similarity 25.18; Pred. No. 4.8e-07;
Matches 115; Conservative 66; Mismatches 165; Indels 112; Gaps 21;

Qy 238 SEIKNP-----AATTDKAQTNTIEPL-----DWDSDSPVQD 269
   ||| | | | - - - - - ||| | | |
Db 9 SQLKNSPLAGNINYEPTWWSRADALKVNENDPPTTQLPSADFPVMDSVTFIINDTPLE 68

Qy 270 VRTGQVANWGYSQLVIAWMGPINQNDNHIIYYL-- --KYGDNELSHWN--- 315
   : | | : | | : | | | | | | | | | | | |
Db 69 L-DGTTVSVNGWSVILLTLTRDRHPNDQP-YLDANGRYDIKRDEWRHGRARMSYWTSRTG 126
   -----VCPFGYNSTAVSQBWSGSAV-LNSDNSIQLFYTRVDFTSDNNTNHQKTASATLYL 369
   : : | | : | | : | | | | | | | | | | | |
Db 127 KDWIFGGYRWAEVSPTREWAGTPIILLNDKGDIDLYTCV-----TPGAIAKVGRGI 180

Qy 370 TDNNGNVSLAQVRNDYIVFEGDGYYIQTVDOMKATNKGDNTAMRD-AHVIEDGSGDRLYL 428
   : : | | : | | | | | | | | | | | | | |

```

```

Db 181 VTSOGVELKDFQVKLFEADQTYIYQTEAQNSSWN-----FRDPSFIPDNDGKLYM 233
Qy 429 VFEASTGLENYOGEDQIYNLWNTGGDAFNKSLFRILSNDIDKSRATWANAAGILKLN 488
Db 234 VFEQ-----NVAGE-----RGSHTVGAAGELGPPVPGHEDVGARFQVCGIG-LAVA 278
Qy 489 KDEKNPKVAELISPLISAPMVSDEIERPNVVKLGKYYLFAATRLNRGSDNDAMWANYA 548
Db 279 KDLGSEB-WEILPPLVAVGVNDQTERPHVYFDGKYYLFTTISH-----KFTYA 326
Qy 549 VG-----DNVAMGVYADSLTSGKPLNDGVLTSVAPNWRATSYVAVPVAGKDDQV 604
Db 327 EGLTGPBGV--YGFVGHFLGPPYRPNMNASGLVL--GNPPEQFQYSHCVMP-----NG 376
Qy 605 LVTSYATNRNGVAGKGM-----STWAPSFLAINDPNT 638
Db 377 LVTSFI---DSVTEGEDYRIGGTEAPTVRILKGDRS 411

RESULT 11
SACB_ACEDI
ID SACB_ACEDI STANDARD; PRT; 584 AA.
AC Q43998;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 01-NAR-2002 (Rel. 41, Last annotation update)
DE Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transferase) (sucrose 6-fructosyl transferase).
GN LSCA.
OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=33996;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 52-61.
RC STRAIN-SRT4 / CBS 550.94;
RX MEDLINE=96253999; PubMed=6704949;
RA Arrieta J., Hernandez L., Coego A., Suarez V., Balmori E.,
Menendez C., Petit-Glatron M.-F., Chamberlain R., Selman-Housein G.;
RT Molecular characterization of the levanucrase gene from the
endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4.;
RL Microbiology 142:1077-1085(1996).
CC -1- FUNCTION: RELEASES FRUCTOOLIGOSACCHARIDES AND LEVAN, A HIGH-
MOLECULAR-MASS FRUCTOSYL POLYMER, FROM SUCROSE. IT ACTS MORE AS A
SUCROSE HYDROLASE THAN AS A FRUCTAN POLYMERASE.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PPM: THE N-TERMINAL IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L41732; AAB36606.1;
DR InterPro; IPR003469; Glyco_hydro.68.
DR Pfam; PF02435; Glyco_hydro.68; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT PROPEP 31 51
FT CHAIN 52 584 LEVANSUCRASE.
SQ SEQUENCE 584 AA; 63525 MW; CBFBE4139AD0B8CE CRC64;

```

Query Match 6.4%; Score 266; DB 1; Length 584;
 Best Local Similarity 26.9%; Pred. No. 1.8e-06;
 Matches 141; Conservative 63; Mismatches 171; Indels 150; Gaps 35;

```

Qy 209 TOMTY---NDFO---KIADTL-IKODGRYTVFFFKASEIKNMPATTK---DAQTNTIEP 258
Db 74 TOQAYDQSDFTARWTRADALQIKAHSDATV-----AAGNSLPAOLTMPTNPADFPVNP 129
Qy 259 -LDVWDSMPVQDVRVQGVANNNGYOLVIAMGIPNQ-----NDNHIY-----LYNKYGDNE 309
Db 130 DVMVWDITWTLIDKHADQF-SYNGMEVIFCLTADPNAGYFDDRHVHARIGFFYRAGIPA 188
Qy 310 LSH-----WKNVGPFGYNSTAV-----SOEWSGSAVLNS--DNSIOLFTRV--- 350
Db 189 SRRPVNGWTYGGHLPDGAQAQYAGYTYNQAEWSSRLMQIHGNTVSVFTYDAFN 248
Qy 351 -DTSNN--TNHQKTASATLYLTDNNGVNSLAQVRNDY--IVFEG-----DGYVYQ 396
Db 249 RDANANNITPPQAITOT-----LGRIHADENHVWFTGTHTAHTPLQPDGVLYQ 297
Qy 397 TYDQWKATNGADN--IAMDRAHVIEDGN--GDRYLVEAST-----GLENYQGEDQIYNW 448
Db 298 -----NGAQNFEENFRDPTTFEDPKHPGVNMYFEGNTAGQGVANCTEAD----- 343
Qy 449 LNYGDDAFNIKSLFRILSNDIDKSRATWANAAGILKLNKDEKNPKVAELISAPM 508
Db 344 LGFRNDP-NAETLOEYLD-----SGAYYOKANIG-LAIATDSTLSK-WKFLSPLISANC 395
Qy 509 VSDEIERPNVVKLGKYYLFAATRLNRGSDNDAMWANYAVG-----DNVAMGVYVADSLT 564
Db 396 VNDQTERPQVYLHNGKYIETISH-----RTTFAAGYVDGPDV--YGFVGDGIR 442
Qy 565 GSYKPLNDGSGVLTASVPANWRTA-----TYSYAVPVAGKDDQVLT 607
Db 443 SDFQPMN-YGSLTGMNPTDLNTAAGTDFDPSDQNPRAFQSYSHVMPGG-----LVE 495
Qy 608 SY---MTNRNGVAGKGMDSWAPSELLQINPONTTIVLAKMTNQ 649
Db 496 SFIDTVENREG-----GTLAPTVRVRI-AQNASAVDLRYNGG 532

RESULT 12
SACB_PSESG
ID SACB_PSESG STANDARD; PRT; 415 AA.
AC 052408;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levanucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
GN LSC.
OS Pseudomonas syringae (pv. glycinea).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PG4180;
RX MEDLINE=98394981; PubMed=3726857;
RT Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
levansucrase genes from the plant pathogens Pseudomonas syringae pv.
glycinea and P. syringae pv. phaseolicola.";
RL Appl. Environ. Microbiol. 64:3180-3187(1998).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```



```
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; AF037443; AAC36056.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
DR KW Transferrase; Glycosyltransferase.
DR SQ SEQUENCE 415 AA; 45844 MW; 839B686AC80610CF CRC64;

Query Match 6.2%; Score 257; DB 1; Length 415;
Best Local Similarity 25.2%; Pred. No. 3.2e-06;
Matches 103; Conservative 66; Mismatches 152; Indels 88; Gaps 19;

QY 261 VWDSPVQDVRTGQVANNNGYOLVIAMMIPNQND-----NHVLL-----YNKYGDNEL 310
DB 44 IWTMPLREL-DGTVVSVNGSVIVTLTADRHDPDPQYVVGANGRVDIKRDWEDRGRARM 102
QY 311 SHWKN-----VGPIFGYNSTAVSOEWSGSAV-LNSDNSTQLFTRVYTSDDNNTNHHK 361
DB 103 CYWYSTRGKDWIFGGRVNAEGVSPTRWAGTPVLLNDKGDIDLYTCV-----TPGAA 156
QY 362 IASATLYLTDNNGVSLAQVRNDYIVFEGDGYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 420
DB 157 IAKVRGRIVTSKGVKELKDFTEVTKFEADKGYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 209
QY 421 DNGDRYLVFEASTGLENYQGEDQIYNWLNLTGGDDAFNIKSLFRILSNDDIKSRATWANA 480
DB 210 PNDGKLYWVFE-----NVAGE-----RGTHVGAAGELGPVPPGHEBTGARGVQV 255
QY 481 AIGIL---KLNKDEKPKVABLYSLISAPMVSDIEIRPNVYKLGNYKYLYFAFATRLNRGS 537
DB 256 CIGLAVAKDLSGDE-----WEILPLVATGVNDOTERPHVFDGKYLYFT-----302
QY 538 NDDAMNANYAGDNV-----AMGYVADSTGSKPLNDSGVLTASVPANWRTATVSY 593
DB 303 -----ISHKFTYADGVTPGPGVGFGEHLFGPYRPMNASGLVL-GNPPAQ-PFQYISHC 355
QY 594 AVPVAGDKDQVLYSYMTNRNGVAGKMD-----STWAPSELLQINDPNT 638
DB 356 VMP-----NGLVTSFI---DSVPTSGEDYRIGTGTEAPTVRILLEGDRS 395

RESULT 13
FNBA_STAAU
ID FNBA_STAAU : STANDARD; PRT: 1018 AA.
AC FI4738;
DT 01-APR-1990 (Rel. 14, Created)
DE 01-APR-1990 (Rel. 14, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Fibronectin-binding protein precursor (FNBP).
GN FNBA
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=8325-4;
RX MEDLINE=8908998; PubMed=2521391;
RA Signaes C., Raucet G., Joensson K., Lindgren P.-E.,
RA Anantharamiah G.M., Hoeck M., Lindberg M.;
RT "Nucleotide sequence of the gene for a fibronectin-binding protein
RT from Staphylococcus aureus: use of this peptide sequence in the
RL synthesis of biologically active peptides."
RL Proc. Natl. Acad. Sci. U.S.A. 86:699-703(1989).
CC -!- FUNCTION: THE ABILITY OF BACTERIA TO BIND FIBRONECTIN HAS BEEN
CC PROPOSED AS A VIRULENCE FACTOR ENABLING BACTERIA TO COLONIZE
CC WOUND TISSUES AND BLOOD CLOTS. BINDING OF PLASMA FIBRONECTIN TO
CC THE BACTERIAL SURFACE MIGHT BLOCK ADHESION RECEPTORS ON S.AUREUS,
CC THUS REPRESENTING AN IMPORTANT DEFENSE MECHANISM AGAINST TISSUE
CC INVASION.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
```

--!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
--!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
--!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 IN THE REGION OF THE MEMBRANE ANCHOR.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL: X59771; CA042442.1; -.
PIR: S15330; FCSOAG.
InterPro: IPR001899; Gram_pos_anchor.
IntraPro: IPR003599; Ig.
Pfam: PF00746; Gram_pos_anchor; 1.
SMART: SM00409; IG; 1.
PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
Cell wall; Transmembrane; Receptor; Repeat; Signal;
Immunoglobulin domain.

SIGNAL 1 37
CHAIN 38 1164 IGA FC RECEPTOR.
DOMAIN 38 1131 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1141 1159 CYTOPLASMIC ANCHOR (POTENTIAL).
FT DOMAIN 1160 1164 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 434 534 IG-LIKE DOMAIN.
FT DOMAIN 199 438 IGA-BINDING (POTENTIAL).
FT DOMAIN 439 826 IGA-BINDING (POTENTIAL).
FT DOMAIN 827 945 PRO-RICH REPEATS.
FT DOMAIN 1131 1137 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 PROTEINS.

SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;

40:

QY	10	VVTLSTAALVFGA-----TTYNAS-----ADTNTENDSST-----	VQV	43
Db	21	VASVAASLFGMSVAHASELVKDDSVKTEVAAPYPSMACTDGCNSSSELETTKMEI	80	
QY	44	TTGDNNDIAVKSVTGLSGOVSAASOTTTIRTS-----NANSASSAANTONSNSQVASSAAIT	99	
Db	81	PTTDLIKKAVEPEKTAGETS--ATDTGKREKOLQOWKKNLKNVDNTTILSHEQKNPEFKTI	139	
QY	100	SSTSAAASL-----NNT-----DSKAAQENTNTAKNDDNQ-----	129	
Db	140	DETNDSDALLENOFNETNRLHLHIKOEHEVEKDKAKQO--KTLKQSDTKVDLSNIDKE	197	
QY	130	---KAAPANESSEAK---NEPAVNVDNSSAAKNDD-----QSSKK-----	164	
Db	198	LNHQKSQVEKMAEQKGITNED---KDSMLKKLIEDTKQAQOAKDKDEAEVKVREELGKL	253	
QY	165	--NTTAKLNKDAENVYK-----AGID---PNSLTDDQIKALNMNFSKAAKSGTOM	211	
Db	254	FSSTKAGLQOEIOEHVKWKTSSBENTQKVDEHYANSLQNAOKSLEELDKATNTNQATOV	313	
QY	212	TYNDF---OKIAD---TLIKQDGRYVVPFFKA-----SEIKNMPAATTKDAQNTNI	256	
Db	314	K-NQFLENAOKLKEIQPLIKET---NVKLYKAMESLEQVEKELKHNSANLEDLVAKSK	369	
QY	257	EPLQVWDS-----N-----WPQDVRTGQVAN	277	
Db	370	EIVREYEGKLNOSKNLPELQULEEASHKLVQVDFRKKFKTSEQVTPKKVRKFDLAAN	429	
QY	278	WNGYQ---LVIAMGIPGNONDHIVLLYKNKYGDNLSHWKNVGPFGYNSTAVGOEWGS	334	
Db	430	ENNOQKIETVSPENITVYVEGEDVAKFTVTAKSCKTT--LDPSDLLTKYPNPSDRISLN	487	
QY	335	AVLNSDNSIQLFYTRVDTSDNNTNHQKTSATLYITDNNGNVSLAQVRNDYIVFEGDGYI	394	

RESULT	14	BAG_STRAG	BAG_STRAG	STANDARD;	PRT; 1164 AA.
ID		BAG_STRAG			
AC		P27951;			
DT		01-AUG-1992 (Rel. 23, Created)			
DT		01-AUG-1992 (Rel. 23, Last sequence update)			
DT		15-JUL-1999 (Rel. 38, Last annotation update)			
DE		19A FC receptor precursor (Beta antigen) (B antigen).			
DE		BAG.			
GN		Streptococcus agalactiae.			
OS		Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC		Streptococcus			
OC		NCBI_TaxID=1311;			
RN		[1]			
RP		SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.			
RP		STRAIN-LA239;			
RX		MEDLINE=911312121; PubMed=1857207;			
RA		Jerlstroem P.G., Chhatwal G.S., Timmis K.N.;			
RT		"The IgA-binding beta antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of binding regions.";			
RT		Mol. Microbiol. 5:843-849(1991).			
RT		[2]			
RN		IDENTIFICATION OF IG-LIKE DOMAIN.			
RP		MEDLINE=97035265; PubMed=8880921;			
RP		Bateman A., Eddy S.R., Chothia C.;			
RT		"Members of the immunoglobulin superfamily in bacteria.";			
RT		Protein Sci. 5:1939-1942(1996).			

```

Db 488 YKTNTDN-----HKAEITIKLNKNSQTVTLKAKDSGNV-----VEKFTTIV----- 533
QY 395 YQTYDQWKATNGADNIAMRDHVE-----DNGDRYLVEASTGLNQGEOOI 445
Db 534 -----QKBEKQVPKTPQKDSKTEKVPQEPKSNKQLOQLIISAQOELEKLE----- 583
QY 446 YNWLNYGDDAFNKSFLRSLNDDIKS-----RATWAN-----AAGILKLNKDE 491
Db 584 -----KAIKELMEQPEIPSPNPEYGIQKSIWESQKEPIQEAITSFKKIIGDS 629
QY 492 KNPKVAELYSPLISAPMYSDEIERP-----NVKLGKHYLYFAATRLNRCNSNDWAWNA 545
Db 630 SSKYYTEHYFNKYSDFMNYQLHAQMEMLTRKVQYMNKYPDNEIKKIFESDMKRTKED 689
QY 546 NYAVGDNAVGVYVADSLTGSYKPL-----NDSGVVLTSVPANMRTATYSYAVPVAGKD 601
Db 690 NYGSLNDALGIFEKYLPTFNKIKQIVDLDKKEVDQQA-----PIPNSEM 739
QY 602 DO-----VLVTSYMTNRNGVAGKGMDSWAPSEFLQINPDNTTTLAKWTNQGDWIDD 655
Db 740 DOAREKAKIAVSKYMS-----KVLGSHQ-----HLQKNNSKIIV-----DLFKEL 780
QY 656 SSENLDMTGLDLSAALGERDKPVDWDLIG-----YGLK-----PHDPATPNDPPTP 702
Db 781 EAIKQOITFDIDNAKTEVDENLVH-DAFSKMNAIVAKFKGLETNTPETPDKPIPELP 839
QY 703 TTPETPTPTPKPKPPENGPOTPTNTPTPEIPLTPETPKPOPEQTNNRLPOT 758
Db 840 QAPDTPQADTPHPESPKEAPRVPESPKTPPEAPHVPSPKAPEAPRVPESPKT 895

RESULT 15
SSP5_STRGN STANDARD; PRT; 1500 AA.
AC P16952; Q54184;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Agglutinin receptor precursor (SSP-5).
GN SSP5 OR SSPB.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE=90236997; PubMed=2185241;
RA Demuth D.R., Golub E.E., Malamud D.;
RT *Streptococcal-host interactions. Structural and functional analysis
RT of a Streptococcus sanguis receptor for a human salivary
RT glycoprotein."
RL J. Biol. Chem. 265:7120-7126(1990).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE=96310377; PubMed=8733238;
RA Demuth D.R., Duan Y., Brooks W., Holmes A.R., McNab R.,
RA Jenkinson H.F.;
RT *Tandem genes encode cell-surface polypeptides SspA and SspB which
RT mediate adhesion of the oral bacterium Streptococcus gordonii to
RT human and bacterial receptors."
RL Mol. Microbiol. 20:403-413(1996).
RN [3]
RP FUNCTION: MAY BIND SALIVARY ACID RESIDUES OF SALIVARY AGGLUTININ
CC (SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH
CC ITS RECEPTOR IN VARIOUS ORAL STREPTOCOCCI MODULATE BACTERIAL
CC COLONIZATION OF ORAL TISSUE AND IS ASSOCIATED WITH REDUCED LEVELS
CC OF DENTAL CARIES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- DOMAINS: THE PR2 REGION, BY SIMILARITY WITH THE PROLINE RICH
CC DOMAINS OF THE S.PYOGENES M6 PROTEIN AND STAPHYLOCOCCAL PROTEIN A,
CC MAY TRAVERSE THE CELL WALL, PEPTIDOGLYCAN AND IS FOLLOWED BY

```

```

CC HYDROPHOBIC AA WHICH MAY FUNCTION TO ANCHOR THE PROTEIN TO THE
CC MEMBRANE.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -1- SIMILARITY: BELONGS TO THE SPAP/SSPS/SPAA FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U40026; AAC44100.1; -
CC PIR: A35186; A35186.
CC InterPro: IPR001899; Gram_pos_anchor.
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC Signal: Repeat; Calcium-binding; Transmembrane.
CC SIGNAL 1 38
CC CHAIN 39 1500 AGGLUTININ RECEPTOR.
CC DOMAIN 39 1474 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 1475 1495 POTENTIAL.
CC DOMAIN 1496 1500 CYTOPLASMIC (POTENTIAL).
CC SIMILAR 164 470 TO M PROTEIN OF S.PYOGENES.
CC DOMAIN 164 470 4 X APPROXIMATE TANDEM REPEATS, HRI.
CC REPEAT 164 241 1.
CC REPEAT 242 323 2.
CC REPEAT 324 405 3.
CC REPEAT 406 470 4.
CC DOMAIN 771 887 3 X APPROXIMATE TANDEM REPEATS, PRI.
CC DOMAIN 1414 1436 PRO-RICH (PR2).
CC DOMAIN 1467 1472 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC PROTEINS.
CC CA_BIND 220 235 POTENTIAL.
CC CA_BIND 301 316 POTENTIAL.
CC CA_BIND 931 950 POTENTIAL.
CC CA_BIND 1300 1315 POTENTIAL.
CC SEQUENCE 1500 AA; 164552 MW; 164552 DCF190E7D44D899F CRC64;

```

```

Query Match 5.6%; Score 229.5; DB 1; Length 1500;
Best Local Similarity 20.7%; Pred. No. 0.0041;
Matches 196; Conservative 120; Mismatches 368; Indels 265; Gaps 45;

QY 3 KSGKNWVYVTLSTAALVFGATTVNASADTNINENNDSTVQVTTGNDIAVKSVTILGSGQV 62
Db 14 KVAKTLGCGAVLGTALIAFADKAV--FADEVTTTSTSTVEVATTGN--PATNLPEAQSEM 69
QY 63 SAASDTTIRTSANASSAANTQNSQVASSAITSSTSSAASLNNT--DSKAAQENTN 121
Db 70 SOVAK---ESQAKAGSKESALPVEYSSADLDKAVADAKSAGVKVYQDETKDKGTATTATD 126
QY 122 TAKNDDTQKAAPANESSEA-----KNEPAVN-----VNDSSAAKNDQSSKK----- 164
Db 127 NAKQDEIKSDYAKQAEIKTTTEAYKKEVAHQAEETDKINAENKAADKYOKDLKSHQE 186
QY 165 -----NTAKLNKDAE-----NVYKAGIDPN-----SLTDDQIKALNKN-- 200
Db 187 EVEKINTANATAKAEYEAQKLAQYQKDLATVKKANEDSQDYQNKLSAVQTELARVQRANA 246
QY 201 -----FSKAAKSGTQMTYNDQFKIADTLTIKODGRYTPFFFKASEIKKNPAATTDAQTNT 255
Db 247 EAKEAYEKAVKENT--AKNEALKYENEAIKQ-----RNETAKATYEAAKQ 290
QY 256 IEPLDVWDSWEPQVDRGTQGVANWN---GYQLVIAMGIPNQNDNHIYLLYNK----- 304
Db 291 YEA-----DLAAIKKANEDNDADYQAKLAAYQTELARVQRANA KAEAYDKAVKENTAKNT 346
QY 305 --YGDNELSHWKNVGPITFGYNSTAVSGWSESVLNSDNLQIPYTRVDTSDNNTNHQKI 362
Db 347 AIQENEAIKQRTAKATYDAAVKVEADLAAYKQAN-----ATNEADYQ-- 392

```

```
QY 363 ASATLYLDNNGNVSLAQRNDYIVFEGDGYYYQTYDOWKATNKGADNIAMRDHVEDG 422
Db 393 AKLAAYQTE-----LARVO-----KANADAKATYEKAVEDNKAKNAIKAENEIEIKOR 440
QY 423 NGDRYLVFEA-----STGLENYQGEQIY-----NWLNYGGDDAFNIKSLF 463
Db 441 NAVAKTDYEAKLAKYEADLAKYKKEFAAYTAALAEAEKSKQDGYLSEPRQSOLNFKSEP 500
QY 464 RILSND-----IKSRATWANAAGILKLNKDEKNPKVAELYSPLIS----- 505
Db 501 NAIRTISSVHOYGOBELDALVKS-----GISPTNDRKKSRAYSYFNAINSNNTY 552
QY 506 APWVSDEIERPNVVKLGNYKYLFAATRLNR-----GSND-----DA 541
Db 553 AKLVLEKRPVDVYTGKLNSSFNGKISKVYVYTYLKETGFNDGTMTMFASDPTVTA 612
QY 542 WNNANYAVGDNVAVGYVAD-----SLTG-----SYKPLNDSGVVLTASVPANWRTAT 589
Db 613 WYN-DYFTSTINNVKVFYDEGQLMNLGTGLVNFSSLNRGNGSGAIDKDAIES---VRN 668
QY 590 YSYVAVPVAGKDDQVLY--TSYMTNRNGVAGKGMSTWAPS-FLLOINPDNTTTLAKMT 646
Db 669 FNGRYPISGSSIKIHENNSAYADSSN--AEKSLGARWNTSEWDTTSSPNNWYGAIVGEI 726
QY 647 NOGDW-----IWDDSSENLDMIGDLSAALPG-----ERDKPVDMDLI--GY 686
Db 727 TOSEISFNWASSKSGNIWFAPNSNINAIGVTPKPVAPTAPQPMYETEKPLEPAPVAPSY 786
QY 687 GLKPHDPA-TPNDPETHPTTP-----ETPNT-ETPNT-KTPKTPENPGTP 726
Db 787 ENEPTPVKTPDOPE-PSKPEEPTYETEKPLEPAPVAPSYENEPTPPVKTPDOPE-PSKP 844
QY 727 QTP-----NTPNT-ETPNT-KTPKTPENPGTP 756
Db 845 EEPNYETEKPLEPAPVAPSYENEPTPPVKTPDOPE-PSKPEEPTYDPLP 892
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 18:23:41 ; Search time 40.2 Seconds
(without alignments)
1885.932 Million cell updates/sec

Title: US-09-995-587A-1

Perfect score: 4129
Sequence: 1 MYKSGKNWAVVTLSTAAALVF.....GMGTLLSMFGLAEINKRRFN 789

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2016.5	48.8	797	2 B28551	levansucrase (EC 2
2	744.5	18.0	473	2 A25040	levansucrase (EC 2
3	741.5	18.0	472	2 J00802	levansucrase (EC 2
4	707	17.1	489	2 F97118	levansucrase [impo
5	356.5	8.6	428	2 H97118	levansucrase [impo
6	299	7.2	423	2 J33771	levansucrase - Zym
7	298.5	7.2	413	2 J03520	beta-fructofuranos
8	297	7.2	423	2 J03519	levansucrase (EC 2
9	291.5	7.1	413	2 S47527	extracellular suc
10	281	6.8	415	2 S39195	levansucrase - Erw
11	261	6.3	1038	2 H90053	hypothetical prote
12	255	6.2	1018	2 A32192	fibronectin-bindin
13	246	6.0	940	2 S19702	IgA Fc receptor pr
14	241	5.8	1164	1 FCSOAG	hypothetical prote
15	237.5	5.8	1134	2 A60234	hypothetical prote
16	231.5	5.6	961	2 B90053	hypothetical prote
17	231.5	5.6	2551	2 B98047	salivary agglutini
18	230	5.6	1661	2 A83695	dehydranase - Strep
19	229.5	5.6	1473	2 A35186	probable peptidogl
20	227	5.5	1337	2 T30291	protein F4N2.10 [i
21	225	5.4	2044	2 A81180	hypothetical prote
22	224	5.4	138	2 D96715	extranase - Strep
23	224	5.4	1072	2 A86827	protein F4N2.10 [i
24	220	5.3	826	2 A45559	sporozoite prote
25	219	5.3	2481	2 D90011	FntB protein [impo
26	215	5.2	1556	2 A60988	saliva-interacting
27	213	5.2	1565	2 S04729	surface antigen pa
28	212	5.1	2817	2 B97033	uncharacterized pr
29	211	5.1	719	2 A42808	Ig light chain-bin

ALIGNMENTS

RESULT 1

B28551

levansucrase (EC 2.4.1.10) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-fructosyl transferase
C:Species: Streptococcus mutans
C>Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
C:Accession: B28551
R:Shiroza, T.; Kuramitsu, H.K.
J. Bacteriol. 170, 810-816, 1988
A:Title: Sequence analysis of the Streptococcus mutans fructosyltransferase gene an
A:Reference number: A91892; MUID:88115184
A:Accession: B28551
A:Molecule type: DNA
A:Residues: 1-797 <SHI>
A:Cross-references: GB:M18954; NID:g153635; PIDN:AAA88584.1; PID:g153636
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 48.8%; Score 2016.5; DB 2; Length 797;
Best Local Similarity 50.4%; Pred. No. 9.1e-97;
Matches 408; Conservative 118; Mismatches 211; Indels 73; Gaps 14;

QY 1 MYKSGKNWAVVTLSTAAALV-FGATTVNASADTNIENNDSTVQVTTGDNDIAVKSVTLGS 59
DB 9 MYKKGKFWVAVITITAMLTGICLSSVQA-----DEANSTQV-----S 45
QY 60 GOVSAASDTTIRTSANASASSAANTQNSQVASSAITSSTSSAASLNNTD--SKAAQ 117
DB 46 SELAERSQVOENTTASSAAENQAETVEOET-----PSTNPAATAVNTDQTKVIT 97
QY 118 ENTNTAKNDTQKAAAPANESSEAKNEPAV---NVNDSSAAKNDQOQSSKKNTAK--LNK 172
DB 98 DNAAVESKASKTKDQAAATVTKTAASTPEVGQTKNEKAKATKADITTPRNTIDEYGLTE 157
QY 173 DAENVVKAGIDPNSLTDDQIKALNMNFSKAASGTMQTYNDFQKIADTLIKQDGRYTV 232
DB 158 QARKIATEAGINLSLTQKQVEALNKVLTSDAQTHQMTQOEFDFKIAQTLIAQDERYAI 217
QY 233 PFFKASEIKNNPAATTAKDAQNTTEPLDWDVSWPQDVRTGOVANWNGYQLVIAAMGIPN 292
DB 218 PYFNAKAIKNKAATTQDAQTGQIADLDVWDSWPQAKTGEVINWNGYQLVYVAMGIPN 277
QY 293 QNDNHIYLLNKKYGDNELSHWKNVGPPIFGYNSTAVSOWSGSAVLNSDSISQIFYYRVD 352
DB 278 TNDNHIYLLNKKYGDNNFDHWKNAGSIFGYNETPLTOWSGSATVNEGSLQIFYTKVD 337
QY 353 SDNNTNHOKIASATLYLTNDNGNVSLAQVRNDYIVF-EG-DGYYYQTYDQWAKNKGADN 410
DB 338 SDXSNNRQIATATVNLGFDQDVRILSVLSEKQVQDVRILSVLSEKQVQDVRILSVLSE 397
QY 411 IANRDAHVIDGNGDRYLVFEASTGQENYQGEQIYNWNLNGGDDAFNFKSLFRILSND 470

30 210.5 5.1 1166 2 T28680 fibrinogen-binding
31 208.5 5.0 1256 1 A43829 muramidase-release
32 208.5 5.0 1463 2 T30290 AAS surface protei
33 206 5.0 1092 2 T30214 fibrinogen-binding
34 205 5.0 1315 2 T28679 fibrinogen-binding
35 202.5 4.9 430 2 JC7379 levansucrase (EC 2
36 202 4.9 1449 2 T30552 glucosyltransferas
37 199 4.8 1093 2 B86748 hypothetical prote
38 199 4.8 1592 2 A38175 glucosyltransferas
39 198.5 4.8 357 2 S21758 glutamic acid-spec
40 197 4.8 877 2 P90070 clumping factor B
41 196.5 4.8 1025 2 S69790 fibronectin-bindin
42 196 4.7 792 2 S70305 hypothetical prote
43 195.5 4.7 1193 2 B96943 2,3-cyclic-nucleot
44 195.5 4.7 1449 2 T30857 glucosyltransferas
45 195 4.7 879 2 S23006 shed acute-phase a

Db 398 IAMRDPHVIEDNGDRYLVEASTCTENYQGEDQIYNFTNYGSSAYNVKSLRFLDQD 457
 QY 471 IKSRTAWANAIGILKLNKDEKNPKVAELYSPLISAPMVSDIEIRPNVVKLGKYYLFAA 530
 Db 458 MYNRASWANAIGILKLGKDKTEVDQFYPLSSWTSWSELERPNVVKLGKYYLFTA 517
 QY 531 TRLNRSGNDAMWANYAYGDNVAMGVYVADSLTGSYKPLKNDSGVLTASVPANRRTATY 590
 Db 518 SRLNNGSNDAMWANYAYGDNVAMGVYVADSLTGSYKPLKNDSGVLTASVPANRRTATY 577
 QY 591 SYIYAVPVAGKDDQVLVTSYMTNRNGVAGKGMDSWAPSEFLQINPDNTTTLAKMTNQG 650
 Db 578 SYIYAVPVAGSDDTLMTAYMTNRNREAVAGKGNSTWAPSEFLQINPDNTTTLAKMTNQG 637
 QY 651 WINDSENLDWIGDLSAALPGRGKPDVMDLI-GYGLKPHDPA--TPNDPPTPTTPE- 706
 Db 638 WINDEPSTTDTGTGLTAYLPGENDGYIDWNVIGGYGLKPHDPA--TPNDPPTPTTPE- 706
 QY 707 -----TPETPNTPKTPKPNPNT-----PNTPEIPL 739
 Db 698 IISPEVDFDGLHVKPVKVNDSAGRIDQSRNCGSLNFAFNVSAAGNISVSKPSKINN 757
 QY 740 TPEPKPQETQNNRLPOTGNANKAMIGL 769
 Db 758 TKTCKAHVSTTEKK-OKKGNSEFAALLAL 786

RESULT 2

A25040
 N:Levansucrase (EC 2.4.1.10) sacB precursor - Bacillus subtilis
 N:Alternate names: sucrose 6-fructosyl-transferase
 C:Species: Bacillus subtilis
 C:Date: 16-Aug-1988 #sequence_revision 20-Feb-1995 #text_change 20-Jun-2000
 C:Accession: S07309; A25040; I39967; H69702
 R:Steinmetz, M.; le Coq, D.; Aymerich, S.; Gonzy-Treboul, G.; Gay, P.
 Moll. Gen. Genet. 200, 220-228, 1985
 A:Title: The DNA sequence of the gene for the secreted Bacillus subtilis enzyme levansucrase
 A:Reference number: S07309; MUID:85295507
 A:Accession: S07309
 A:Molecule type: DNA
 A:Residues: 1-473 <RES>
 A:Cross-references: EMBL:X02730; NID:g40118; PIDN:CAA26513.1; PID:g732568
 R:Experimental source: Marburg
 R:Shimotsu, H.; Henner, D.J.
 J. Bacteriol. 168, 380-388, 1986
 A:Title: Modulation of Bacillus subtilis levansucrase gene expression by sucrose and raffinose
 A:Reference number: A25040; MUID:87008406
 A:Accession: A25040
 A:Molecule type: DNA
 A:Residues: 1-68 <SHI>
 A:Cross-references: GB:M14202; NID:g143485; PIDN:AAA22725.1; PID:g143486
 R:Fouet, A.; Arnaud, M.; Klier, A.; Rapoport, G.
 Biochem. Biophys. Res. Commun. 119, 795-800, 1984
 A:Title: Characterization of the precursor form of the exocellular levansucrase from Bacillus subtilis
 A:Reference number: I39967; MUID:841178454
 A:Accession: I39967
 A:Molecule type: DNA
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-11, 'I', 13-62 <RES>
 A:Cross-references: GB:X01987; NID:g143483; PIDN:AAA22724.1; PID:g143484
 R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Broutlet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
 tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 y, M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
 A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, M.

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A:Reference number: A63580; MUID:98044033
 A:Accession: H69702
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-473 <KUN>
 A:Cross-references: GB:299121; GB:AL009126; NID:g2635827; PIDN:CAB15450.1; PID:g26358
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: sacB
 C:Keywords: extracellular protein; glycosyltransferase; hexosyltransferase
 F:1-29/Domain: signal sequence #status predicted <SIG>
 F:30-473/Product: levansucrase #status predicted <MAT>

Query Match 18.0%; Score 744; DB 2; Length 473;
 Best Local Similarity 36.8%; Pred. No. 2.6e-31;
 Matches 183; Conservative 87; Mismatches 151; Indels 76; Gaps 17;
 QY 199 MNFSAKSGTOMTY-----NDFQKTADTLIK 225
 Db 1 MNKFKAKQATVLTFTTALLAGGATQAFKTNQPKYKTYGISHIRHDMQLQPEQ--Q 58
 QY 226 ODGRYTVPEFKASEIKNMPAATTDAQTNTIEPLDWDSPVQDVRTGQVANNNGYQLVI 285
 Db 59 KNEKYQVPEFDSSTIKNISSA-----KGLDWDSPQLONA-DGTVANYHGHVIF 107
 QY 286 AMMGIP-NONDNHLYLYNKKYGDNELSHKKNVGP-----FGYNSTAV---SQWSSGA 335
 Db 108 ALAGDPKNADTSYMFYQKVGTSIDSWNAGRVKDSKDFANDSILKDKQWSSGA 167
 QY 336 VLNSDNTSTLPYTRVDSDNNNTNOKTASATYLTLDNNGNVSLAQVRNDYIVFEGDGY 395
 Db 168 TFTSDGKIRLYFT--DFSGKHGKQTLTAQVNVSSDSSNLINGVEDYKSIFFDGDKTY 225
 QY 396 QTYDQW--KATNGADNIAHMDAHVIDGNGDRLVFEASTGLEN-YOGEDQIYNWLN 452
 Db 226 QNVQFIDEGNYSNGDNHRLDRPHYVED-KGHKLYLFEANTGTEDYOGESLFRNKAYG 284
 QY 453 GDADFNIKSLFRILSNDIDKSRATWANAALGILKLNKDEKNPKVAELYSPLISAPMVSD 512
 Db 285 KTSFFQESQKLLQSDK-KRTAELANGALGMIELNDYTLKKVMK---PLIASNTVTDE 340
 QY 513 IERNVVKLKNKYLFAATRLNRGSDNDAMWANYAYGDNVAMGVYVADSLGSKYKPLND 572
 Db 341 IERANFKMNGKWLFTD---SRGSK---MTIDGITSNDIYMLGIVNSLTGPKPLNK 393
 QY 573 SGVVLTAASPANRRTATYSYAVPVAGKDDQVLVTSYMTNRNGVAGKGMDSWAPSEFLQ 632
 Db 394 TGLVLMKDLDPNDVTYTSHEAVPOA-KGNVNVITSYMTNRGFIADK--QSTFAFSFLN 450
 QY 633 INPDNTTTLAKMTNQG 649
 Db 451 IKGKTSVVKDSILEQG 467

RESULT 3

J00802
 N:Levansucrase (EC 2.4.1.10) precursor - Bacillus amyloliquefaciens
 N:Alternate names: 2,6-beta-D-fructan-6-beta-D-fructosyltransferase; sucrose
 C:Species: Bacillus amyloliquefaciens
 C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 15-Oct-1999
 C:Accession: J00802; S11739
 R:Tang, L.B.; Lenstra, R.; Borchert, T.V.; Nagarajan, V.
 Gene 96, 89-93, 1990
 A:Title: Isolation and characterization of levansucrase-encoding gene from Bacillus
 A:Reference number: J00802; MUID:91092506
 A:Accession: J00802
 A:Molecule type: DNA
 A:Residues: 1-472 <TAN>
 A:Cross-references: EMBL:X52988; NID:g39333; PIDN:CAA37179.1; PID:g39334
 A:Experimental source: ATCC 23844

Db	57	NP	NEFKLTAPNLVWDTWPLVK	-KDGSLAVVVGKYVIFALTASRNVGNMKRHVAGISYFC	115
QY	307	DNEJSHKWNKGPFGYNSTAVSGEWSGAVLNSDSNSIOLEFYTRVDTSDN	-NTNHQKIAS	364	
Db	116	STDGDNWVYKGLAYNVEDALGSRQWAGSAILDENGWQFFTTATGRKGEAVRFEQLVK	175		

C/Genetics:
A;Gene: CAC1774

Ob 116 STDGENWYKGLAYNYVEDALGSRQWAGSAILDENGWQFFYATCRKGEAVRTFEQLVK 175

[illegible]

```

257 QY EPDLVDWDSHPQDVRTGQVA--NWNGYQLVIAMGMIPNODNH---TYLLNKKYGDNEL 310
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
38   Db DLWLWDLTWPLRLDINGNPVSFKGWNVIFSLVADRNP-WNDRSHARGYFYFSKDGKS-- 94
    QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
311 QY SHMKNGPIFGYNSTAVSGSEGS AVL--NSDNSQLFTYTRVDTSNNNTNHQKIASAT-- 366
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
95   Db -WYVGCHLLQELANRTAEWSGGTIMAPGSRNOYETFTST-LFDKNGVREAAVAATKG 151
    QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
367 QY -LYLTDNNGNVSLAQVRNDYIIVFEGDGYGYQYDQWKATNGKADNIARDAHV-IEDGNG 424
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
152 QY RIV-ADSEG-VNFKGFDQSTDLFOADGLFYQNYAENLWN-----FRDPHFVFINPDG 202
    QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 QY DRYLVFEASTGLENYQGEDQYIYNWLNUNYGGDDAFNTKSLFRILSNDIDKSRATWANAATGI 484
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 QY EYALFEAN--VATVRGDEDI-----GED-----EIGPVPAVTVPKDALICSASIGI 248
    QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
485 QY LKLNKDEKNPKVAEYLSPLISAPMVSDEIERPNVVKLGNYKYLFAAATRLNGSDNDAMWN 544
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
249 QY ARCLSPDRTE--WELLPLLTAFGVNDQWQMRPHVIFONGLYLETFISH-----D 295
    QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
545 QY ANYAVG--DNVAMVGYVADS-LTGSYKPLNDSGGVLTASVPANWRTATYSYIAPVPVAGKD 601
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
296 QY STYADGLTSGDGLYGFVSENGIFGPEPLNGSGVLVG--PASOPTYAYAHYIM----- 347
    QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
602 QY DOVLVTSYMTN---RNG--VAGKMDSTWAPSLFLOINPDNT 638
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
348 QY NGLVSEFINEIIDPKSGKVTAGSL-----APTVRVLOGHET 386
    Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8
JC2519
levansucrase (EC 2.4.1.10) precursor - Zymomonas mobilis
N:Alternate names: sucrose 6-fructosyltransferase
C:Species: Zymomonas mobilis
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text

C:Accession: JC2519; PC2376
R:Kyono, K.; Yanase, H.; Tomomura, K.; Kawasaki, H.; Sakai, T.
Biosci. Biotechnol. Biochem. 59, 289-293, 1995
A:Title: Cloning and characterization of *Zymomonas mobilis* genes encoding extracellular
A:Reference number: JC2519; MUID:95218269
A:Accession: JC2519
A:Molecule type: DNA
A:Residues: 1-423 <KYO>
A:Cross-references: DDBJ:D17524; NID:9809529; PIDN:BAA04475.1; PID:9809530
A:Accession: PC2376
A:Molecule type: Protein
A:Residues: 1-19 <KY2>
C:Genetics:
C:Gene: sucZE2
C:Keywords: glycosyltransferase; hexosyltransferase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-423/Product: levansucrase #status predicted <MAT>

Query Match 7.2%; Score 297; DB 2; Length 423;
Best Local Similarity 26.4%; Pred. No. 2.9e-08;
Matches 119; Conservative 75; Mismatches 155; Indels 102; Gaps 23;
QY 250 DAQNTTIEPLD-----VMSWVQDVRTGOVANNNGYOLVIAMM-----GIPN 292
DB 24 DDPTATPTIDYDFPMTDKYVWDTWPLRDI--NGQVVSFGWVIFALVADRKYGWHN 82
QY 293 QNDN-HIYLLNKKYGDNELSHWKNVGFYGNSTAVSQWSSGSAVL--NSDNTSLFYTR 349
DB 83 RNDGARIGYFYSRGSN----WIFGGHLLKDGANPRSNWSGGCTIMAGTANSVEVFTS 138
QY 350 VDTSDNTNHOI-KT-ASATYLTLDNNGNVSQAQRNDYIVFEGDGYGYQTYDQWKNATKGA 408
DB 139 V-----NDTPSESVPACQKGIYADKSWFDFGDKVTDLFOADGLYADI-----AENFW 190
QY 409 DNTAMRDAHV-IEDNGDRYLVEASTGLNGYQDQIYNWLYGGDDAFNIKSLFRILS 467
DB 191 D---FRDPHFVFNPEDEGKTVALEGVNVMOR-----GAVAVGEEIEGVPVP 233
QY 468 NDIKSRATWANAAGILK-LNKDEKNPKVAELYSPLISAPMVSDIETERNPVVKNKYY 526
DB 234 KTTPTGARYAAAGIAQALNEARTEWK---LLPLVTAFGVNDQTERPHVVFQNGLY 290
QY 527 LFAATRLNGSNDNANANYAVG----DNVAMVGVVADS-LTGSYKPLNDSGVLTASV 581
DB 291 LFTISH-----HSTYADGLSGPCV--YGFVSENGIFGPEPLNGSLVL--GN 335
QY 582 PANWRTATYSYAVPVAGKDDQVLTYSMTN-----RNGVAGKGMDSWAPSFLQ 632
DB 336 PSSQPTQAYSHYVM-----TNGLVTSFDTIPSSDPNRYRG-----GTLAPTVKLE 382
QY 633 INPDNTTTLAK-----MTNQGDIWDDSEN 659
DB 383 LVGHRSFVTEVKGYGYPQIEWLAEDSSN 413

RESULT 9
S47527
extracellular sucrose - *Zymomonas mobilis*
C:Species: *Zymomonas mobilis*
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-1999
C:Accession: S47527
R:Song, K.B.; Lee, S.K.; Joo, H.K.; Rhee, S.K.
Biochim. Biophys. Acta 1219, 163-166, 1994
A:Title: Nucleotide and derived amino acid sequences of an extracellular sucrose gene (s
A:Reference number: S47527; MUID:94368848
A:Accession: S47527
A:Molecule type: DNA
A:Residues: 1-413 <SON>
A:Cross-references: EMBL:L08094

Query Match 7.1%; Score 291.5; DB 2; Length 413;
Best Local Similarity 26.4%; Pred. No. 5.3e-08;
Matches 106; Conservative 77; Mismatches 147; Indels 71; Gaps 22;
QY 257 EPLDVWDSWVQDVRTGOVANNNGYOLVIAMMGPNDNH-----IYLLNKKYGDNEL 310
DB 38 DDLWLDWTPURDINGNPVSFGWNVIFSVFADRNIP-WNDRHSHARIGYFYSKDGKS-- 94
QY 311 SHWKNVGFYGNSTAVSQWSSGSAVL--NSDNTSLFYTRVDTSDNNTNHOIATASAT-- 366
DB 95 --WVYGHLLQESANTRTAESGGTIMAPGSRNQVETFTST-LFDKNGVREAAVATKG 151
QY 367 -LYLTLDNNGNVSQAQRNDYIVFEGDGYGYQTYDQWKNATKGMDSWAPSFLQ 424
DB 152 RIY-ADSEG-VWFKGFDQSTDLFOADGLFYQNAENLRN-----FRDPHFVFNPEDEG 202
QY 425 DRYLVFEASTGLNGYQDQIYNWLYGGDDAFNIKSLFRILSNDIDIKSRATWANAAGI 484
DB 203 ETYALFEAN--VATVRGEDI-----GED-----EIGVPANTVVPKDNLCSSASIGI 248
QY 485 LKLNKDEKNPKVAELYSPLISAPMVSDIETERNPVVKNKYYLFAATRLNGSNDNANW 544
DB 249 ARCLSPDRT--WELGPPLLTAFGVNDQMERPHVIFQNGLYLFTISH-----D 295
QY 545 ANYAVG--DNVAMVGVVADS-LTGSYKPLNDSGVLTASVPANWRTATYSYAVPVAGKD 601
DB 296 STADGLTSGDGLYFVSENGIFGPEPLNGYGLVGG--PASQPTAVAHYIM-----347
QY 602 DQVLTYSMTN-----RNGVAGKGMDSWAPSFLQINPDNT 638
DB 348 NNGLVESFINEIIDPKSGKVRAG--GSLAPTVERVLEQGHET 386

RESULT 10
S39195
levansucrase - *Erwinia amylovora*
C:Species: *Erwinia amylovora*
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: S39195
R:Geier, G.; Gelder, K.K.
submitted to the EMBL Data Library, September 1993
A:Description: Characterization and influence on virulence of the levansucrase gene
A:Reference number: S39195
A:Accession: S39195
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-415 <GEI>
A:Cross-references: EMBL:X75079; NID:9433558; PIDN:CAA52972.1; PID:9433559

Query Match 6.8%; Score 281; DB 2; Length 415;
Best Local Similarity 25.6%; Pred. No. 1.9e-07;
Matches 109; Conservative 73; Mismatches 137; Indels 106; Gaps 22;
QY 257 EPLDVWDSWVQDVRTGOVANNNGYOLVIAMMGPNDNH-----QNDNHIYLL-----YNYKG 305
DB 40 EEVFIWDTMPLRDF-DGEIISVNGWCIIFTLTADRTDNFQFDENGNYDITRDWEDRHG 98
QY 307 DNELSHW-KNVGP--IFGYNSTA-----VSOEWSGSVAV-LNSDNTSLFYTRVDTSDNNT 357
DB 99 RARICYWYRTGKDWIFGGRVMAEGVAPTTREWAGTPTILLNDRGDDIDLYITCV-----T 152
QY 358 NHOKIASATLYLTDNNGNVSQAQRNDYIVFEGDGYGYQTYDQWKNATKGMDSWAPSFLQ 414
DB 153 PGATIAKVRKIVTSQVSLGFGQVTSLSFADGTIYQTEQNAFWN-----FR 202
QY 415 D-AHVIEDGNGDRYLVEASTGLNGYQDQIYNWLYGGDDAFNIKSLFR 464
DB 203 DPSPFIDRNDGKLYMLFEGN--VAGPRGSHETQAEWNVPPGVEDVGG-----249
QY 465 ILSNDDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPMVSDIETERNPVVKNK 524
DB 250 -----AKYQAGCVG-LAVAKOLSGSE-WQIPLPITAVGVNDQTERHFEVFDQK 297

```

Query Match          6.3%; Score 261; DB 2; Length 1038;
Best Local Similarity 18.2%; Pred. No. 7.3e-06;
Matches 199; Conservative 151; Mismatches 301; Indels 442; Gaps
52;

QY 55 VTLSGQVSAASD---TTITSANASASSAANTQNSOV-----ASSAAITSSSTSS 104
Db 27 VGMGDKEAAEQKTTTVEENGNSADNKTSETQTTATNVNHIETQSYNATVTEQPSN 86
QY 105 AASLNNITDS-KAAQ-----ENTINTAKND-----TOKAAPA 134
Db 87 ATQVTTTEAPRAVQAPQTAQPAWVETKKEEPQVKETTPQDNSGNQRVDLTPKKVYQ 146
QY 135 NESSE-----AKNEPAVNVN-DSSAKNDQDOSSKNT--TAKLNKDENVYKK 180
Db 147 NGGTQVEVAQPRTASESPRVTRSADVAEAKESADYSEVKGTDVTSKVTVESGSEAP 206
QY 181 AG--IDPNSLTDDQIKALKNWFSKAAKSGTOWTNDFOKIADTLLIKODGRYTVPPFF-KA 237
Db 207 CGNKVEPHA--GORVVLKYKLAFADGLKRG---DYDFD-----TILSNVNVTGYSTARKV 256
QY 238 SEIKN-----WPAATTKDAQTNTIE-PLDVMDSWPVQ---DVRTGOVANWNGYQL 283
Db 257 PEIKNGSVVMATGEILGNIRYTFTENEHKEVETANLEINLFDIPRTVQS---NGEQK 313
QY 284 VIANM-----GIPNQ--NDNHIYLLYNKYGDNELSHWKNGYGPFGYNSTA 326
Db 314 ITSKLNGEETKEIPVYVNGVSNYSYTVNGVSIETFNK-ESNKFTTHIAYIKPNNGQSNY 372
QY 327 VSQWES---GSAVLNSDNSIQLF-----YTRVDTSNN--TNHQKIASATLYLT 370
Db 373 VSYVTGLTEGNSLAGQPTVKYVEYLVKKDLPOSVTANTSTFNKFDVTKEMNGKLSVQ 432
QY 371 DNNG-NVSLAQRYNDYIYFEGDGYIYOTYDOWKATKNKADNIAHMDAHDVEDGNGDRYL 429
Db 433 DNGSYSLNDKLDKTYVI-----HYTGEYLQ-----GSDQVNER----- 466
QY 430 FEASTGLENYQGEDQIYNWNLVGGDDAFNKLSPRLNSDDTKSRATWANAIGTLKLNK 489

```


RESULT 15
A60234

IgA FC receptor precursor - Streptococcus agalactiae (strain SB35)
N:Alternate names: IgA-binding protein; protein Bac
N:Contains: beta antigen
C:Species: Streptococcus agalactiae
C:date: 08-Dec-1992 #sequence.revision 08-Dec-1992 #text_change 26-Aug-1999
C:C-accession: A60234; S14595;
R:Heden, L.O.; Frithz, E.; Lindahl, G.
Eur. J. Immunol. 21, 1481-1490, 1991

A:title: Molecular characterization of an IgA receptor from group B streptococci: se e
e:ents with IgA-binding capacity.
A:reference number: A60234; MUID:91257158
A:A-accession: A60234
A:A:molecule type: DNA
A:A:Residues: 1-1134 <HE2>
A:A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CRAA11384.1; PID:g46521
A:A:Note: the source is designated as group B streptococcus strain SB35
Submitted to the EMBL Data Library, G
frgments.
A:Accession: Molecular characterization of an IgA receptor from group B streptococ
A:Reference number: S14595
A:Accession: S14595
A:Molecule type: DNA
A:Residues: 1-1134 <HE2>
A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CRAA11384.1; PID:g46521
A:Note: the source is designated as Streptococcus agalactiae
Lindahl, G.; Akerstroem, B.; Vaerman, J.P.; Stenberg, L.
Eur. J. Immunol. 20, 2241-2247, 1990

A:Title: Characterization of an IgA receptor from group B streptococci: specificity for
A:Reference number: A60230; MUID:91055597
A:Accession: A60230
A:Molecule type: protein
A:Residues: 'X',39-48,'X',50-52,'X',54-56 <LIN>
C:Superfamily: IgA Fc receptor
C:Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein
F:1-37/Domain: signal sequence #status predicted <SIG>
F:38-1134/Product: IgA Fc receptor #status experimental <MAT>
F:139-438/Domain: IgA binding #status predicted <IGAI>
F:439-826/Domain: IgA binding #status predicted <IGAI>
F:827-915/Region: proline-rich repeats
F:916-1101/Domain: cell wall-spanning #status predicted <CWS>
F:1102-1129/Domain: transmembrane #status predicted <TM>

Query Match 5.8%; Score 237.5; DB 2; Length 1134;
Best Local Similarity 20.1%; Pred. No. 0.00014;
Matches 189; Conservative 129; Mismatches 349; Indels 273; Gaps 39;

QY	6	KMAVVTSTAALVFGATTVNASADTNINNDST-----VOVTTGDNDAIAKSVTLGS	59
DB	42	KDSVKTTEVAAPY-----PSMAQTDQGNSSSELETTKMEIPTDIKKAVERPEKTA	96
QY	60	GOVSAASDTTIRSA-----NANSASSAANTONSQVASSAITSSTSSAASL-----	108
DB	97	GETS-ATDTGKREKQLQWKNLKNVDNTILSHQKNEFKTKIDETNDSALLELENOF	155
QY	109	NNT-----DSKAAQENTAKNDTQ-----KAAPANESSEAK- 141	
DB	156	NETNRLHLIKOHEVEKOKKAKQ--KTLKQSDTKVDLSNIDKELNHQKSOVERKMAEQK	213
QY	142	--NEPAVNVNDSAAKNDD-----QSSKK-----NTAKLNKDAENV	178
DB	214	INED-----KSMLKAKIEDIRKQAQADKEDAEVKVREELKGLFSKAGLQOEIOEHV	269
QY	179	KK-----AGID-----PNSLTDQIKALKNFNSKAAKSGTQMTYNDF-----OKIAD--	221
DB	270	KKETSSEENTQVDHYANSQNLAKSLELDKATTEQATQYK--NOFLENAQKLKEIQ	328
QY	222	TLIKODGRYVFPFKA-----SEIKNMPAATTKDAQTNTIEPLDVNDS-----	264
DB	329	PLIKET-----NVKLYKAMESLEQVEKELKHNSANLEDLVAKSEIVREYEGKLNQSKNL	385
QY	265	-----WPVQDVTCQVANWNGYQ-----LVIAMMGI 290	
DB	386	PELKQLEEAHSLKQVVEDEFKFKTSEQVTPKKVRKDRDLAANNENNOOKIELTVSPENI	445
QY	291	PNQNDNHLYLLYKNGNLSHKNVGPFGYNSTAVSQWGSVAVLNSDINSIQLFYTRV	350
DB	446	TVYEGEDVKTVTAKSDSKTT--LDFSDLITKYNPSVSDRISTNYKTNTDN-----HKIA	498
QY	351	DTSDNTNHQIKASATLYLTDNNGVSLAQVRNDYIVFEGDGYIYQTYDQWKATNKADN	410
DB	499	EITIKNLKLNESQTVLKAQDDSGNV-----VEKFTITV-----OKKEEQVPEKT	544
QY	411	IAMRDAHVLE-----DGNDRYLVFEASTGLNKGEDQIYNMLNYGGDDAFNIKS	461
DB	545	PEQKDSKTEKVPQEPKSNKQNLQELIKSAQOEKLE-----KA 585	
QY	462	LFRILSNDIKS-----RATWAN-----AAGILKLNKDEKNPKVABEYSP LISAP	507
DB	586	IKELMEQPEIPSNPEVGIQSWESQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSD	645
QY	508	MVSDTERP-----NVYKLGKNKYLFAATRLNRGSDNDAMNANYAVGDNVAMGYVAD	561
DB	645	FMNYQLHAQEMLETKRVVQYNNKYPDNAEIKKIFESDMKRTKEDNYGSLNDALKGYFEK	705
QY	562	SLTGSYKPL-----NDSGVVLTAASVPANWRTATYSYAVPVAGKDDQ-----VLVTSYMT	611
DB	706	YELTPFNKIKQIVDDLKKVQDQPA-----PIPENSEMDOAKAKIAVSKYNS	755
QY	612	NRNGVAGKGMDSWNAPSFLQLQINPDNTTTLAKMTNOGDWIDDSSENLDIGDLSAAL	671

Search completed: September 26, 2002, 18:25:06
Job time: 85 sec

Db	756	-----KVLGQVHQ-----HLQKKNHKIV-----DLFKELEAIKQOITFIDIDNAKT	796
QY	672	PGERDKPVDWDLIG-----YGLK-----PHDRATPNDPPTPTTPTPTPTPTPTPK	718
Db	797	EVEIDNLVH-DAFSKMNATVAKFQGLENTPTPTDTTKIPELFOAPDTPOADTPHVE	855
QY	719	TPENPCTPQTPATPT	758
Db	856	SPKAPAPRVPSPKTPEAPHVPSPKTPEAPKIPPEPKT	895

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 18:23:41 ; Search time 61.92 Seconds
(without alignments)
1415.330 Million cell updates/sec

Title: US-09-995-587A-1

Perfect score: 4129

Sequence: 1 MYKSGKNWAVVTLSTAALVF.....GMGTLLSMFGLAEINKRRFN 789

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	798	19.3	881	22	ABG21651
2	798	19.3	893	22	ABG07167
3	793.5	19.2	1027	22	ABG21645
4	766	18.6	736	22	ABG22589
5	760.5	18.4	774	22	ABG05590
6	760.5	18.4	774	22	ABG11754
7	759.5	18.4	789	22	ABG04946
8	759.5	18.4	789	22	ABG28407
9	759.5	18.4	1095	22	ABG21602
10	759.5	18.4	1095	22	ABG21823
11	758.5	18.4	527	22	ABG04307

12	758.5	18.4	789	22	ABG23205	Novel human diago
13	758.5	18.4	823	22	ABG25769	Novel human diago
14	756.5	18.3	493	22	ABG21647	Novel human diago
15	756	18.3	546	22	ABG21832	Novel human diago
16	756	18.3	562	22	ABG21605	Novel human diago
17	754.5	18.3	855	22	ABG04500	Novel human diago
18	749	18.1	929	22	ABG23199	Novel human diago
19	744	18.0	644	22	ABG06230	Novel human diago
20	744	18.0	913	22	ABG02461	Novel human diago
21	741.5	18.0	472	12	AA10671	B.amyloidetfacien
22	739.5	17.9	797	22	ABG18136	Novel human diago
23	736	17.8	689	22	ABG00041	Novel human diago
24	732	17.7	1551	22	ABG28409	Novel human diago
25	729.5	17.7	786	22	ABG02257	Novel human diago
26	718	17.4	542	22	ABG21613	Novel human diago
27	679.5	16.5	436	22	ABG04298	Novel human diago
28	671.5	16.3	657	22	ABG21612	Novel human diago
29	671.5	16.3	657	22	ABG21834	Novel human diago
30	659	16.0	3048	22	ABG25791	Novel human diago
31	656	15.9	778	22	ABG13422	Novel human diago
32	648.5	15.7	828	22	ABG24549	Novel human diago
33	648.5	15.7	928	22	ABG21821	Novel human diago
34	648.5	15.7	928	22	ABG22394	Novel human diago
35	648.5	15.7	928	22	ABG26921	Novel human diago
36	646.5	15.7	914	22	ABG09624	Novel human diago
37	646	15.6	708	22	ABG23779	Novel human diago
38	646	15.6	708	22	ABG28401	Novel human diago
39	634.5	15.4	495	22	ABG04320	Novel human diago
40	626	15.2	893	22	ABG26977	Novel human diago
41	617.5	15.0	426	22	ABG04288	Novel human diago
42	615	14.9	405	22	ABG21859	Novel human diago
43	613.5	14.9	487	19	AAW48306	Bacillus sp. V230
44	611	14.8	692	22	ABG21599	Novel human diago
45	603	14.6	417	22	ABG04289	Novel human diago

ALIGNMENTS

RESULT 1
ABG21651
ID ABG21651 standard; Protein; 881 AA.
AC ABG21651;
XX
XX
DT 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #21642.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX
XX WO200175067-A2.

XX
XX 11-OCT-2001.

XX
XX 30-MAR-2001; 2001WO-US08631.

XX
XX 31-MAR-2000; 2000US-0540217.

XX
XX 23-APR-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

XX
XX Drmanac RT, Liu C, Tang YT;

XX
XX WPI; 2001-639362/73.

XX
XX N-PSDB; AAS85838.

XX
XX New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess

QY	561	DSLTSYKFLPNDGGVLTASVPANWRTATSYVAVPAGKDDQVLVTSYMNRRGVAGKG	620
		: : :	
Db	791	nsaltgypkplnktglvlqmgldpndvtfysnfavpqa - kgnvvitsymtnrgfedk -	848
		: : :	
QY	621	MDSTWAPSFLLQINPDNTTTLAKMTNQG	649
Db	849	-katrapsflmniknktsvvksnilegg	876
		: : :	
		-katrapsflmniknktsvvksnilegg	
RESULT	2		
ABG07167			
ID	ABG07167	standard; Protein; 893 AA.	
XX			
AC	ABG07167;		
XX			
DT			
XX	13-FEB-2002	(first entry)	
DE			
XX	Novel human diagnostic protein #7158.		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KX	food supplement; medical imaging; diagnostic; genetic disorder.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			

30-MAR-2001; 2001WO-US08631.
31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HYSE-) HYSEQ INC.
Drmanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
N-PSDB; AAS71354.
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity
Claim 20; SEQ ID NO 37526; 103pp; English.
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
Sequence 893 AA;

Query Match 19.3%; Score 798; DB 22; Length 893;

[illegible][illegible]

PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
XX	biodiversity.
PS	Claim 20; SEQ ID NO 35305; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting and
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 789 AA;
	Query Match 18.4%; Score 759.5; DB 22; Length 789;
	Best Local Similarity 37.2%; Pred. No. 9.7e-39;
	Matches 193; Conservative 90; Mismatches 187; Indels 49; Gaps 18
QY	150 DSSAAKNDDQQSKNTTAKLNKDAENVVKKAGI---DPSNITDDQIKALNMNFNSKA 206
DB	: : : : : : : : : : : : : : : : : : : :
DB	296 dsirednrgsrkthvhtegdmnnlkkivkgatvitftailagatqafakennqayk 355
QY	207 SG---TQMYYNDFOKIADTLIKODGRYTVPFFKASEIKNPMAATTKDAQTNIEPLDVWD 263
DB	: : : : : : : : : : : : : : : : : : : : : : : : :
DB	356 etygvshtrhdmlqlpkq--gqnkyqvpgfdqstknies-----kgldvwd 403
QY	264 SWPQDVRTGTGVANNWGYQLVTAMMGIPI-NQDNHIIYLKYNGDNEUSHKNVGPI--- 319
DB	: : : : : : : : : : : : : : : : : : : : : : : : :
DB	404 swplqna-dglvaeyngyhvfalagspkdaddtslymfykvgdsidswnagrvmfkd 462
QY	320 --PCYNSTAV---SOEWSGSVAIVNSDSNIOLFTRYTRVDTSNNTHQKIASATLYLTENN 373
DB	: : : : : : : : : : : : : : : : : : : : : : : : :
DB	463 sdkfaodnpilddqtqewsgsatfsdgkrlfyt--dysgkhgykgslltaqvnvkskd 520
QY	374 GNVSIAQVRNDYIVFEGDGYYYIQFYDOW--KATNKAGDNTAMRDHAHVIEDGNDRYLVEF 431
DB	: : : : : : : : : : : : : : : : : : : : : : : : :
DB	521 dtlikingvedhktifdgdgkytnqqvqfidegnvtsgdnthtirdphyyed-kghkylyve 579
QY	432 ASTGLEN-YOGEQIYNWLNYGGDDAFNIKSLFRILSNDDIKSRATWANAIGITLKLND 490
DB	: : : : : : : : : : : : : : : : : : : : : : : : :
DB	580 antgtengyggeeslnfxayyygggnffrkesqk-lqqsakkdaelangalglielnd 638
QY	491 EKNPKVAELYSPLISAPMWSDIERPNVVKLGNKYLFYAATRLNRSGNDWMNANYAVG 550
DB	: : : : : : : : : : : : : : : : : : : : : : : : :
DB	639 ytlkkvmk---plltantvtdeieranfvkmngkwylfdrsgrskmtldg--lnsn---- 689
QY	551 DNVMYGVYVADSUTGSYKPLNDSCVVLTVASVPANWTATSYAYVAVPAKDQDLVTSYM 610
DB	:: : : : : : : : : : : : : : : : : : : : : : : : :
DB	690 -diymlygvysnltpgypklntglvlgmglpdndvftyshfavpqa-kgnnvvtitsym 747
QY	611 TWRNGVAGKMDSWPSPFLQINPDNTTTLVLAKTNQ 649
DB	: : : : : : : : : : : : : : : : : : : : : : : : :
DB	748 trngffedk--kattapsflmkanktksvtknsleaa 794

Db	356	etygvshitrhdmqlipkq--qqnekyvqpqfdqstiknlesa-----kgldvwd 403
QY	264	SWPQDVRTGOVANNWGQYLVIAMGIP-NONDNIHYLLYKNGDNLSHWKNVGPI--- 319
Db	404	swplqna-dgtvaeingvhwvfallagspkdaddtslymfyqkgvqgdnswknaagrvmkd 462
QY	320	---FGYNSTAV---SQEWSGSAVLNDSNLSIOLEFTRVDTSNDNTNHOKIASATLYLTDNN 373
Db	463	sdkfandpilkdqeqwsgsatfsgdkirlyf--dysgkhygkqslttacqvnvsksd 520
QY	374	GNVSLAQRNDYIVFEGDGYGYQYVDQM--KATNKGADNIAMRDAHVIEDNGDRYLVE 431
Db	521	dtikngvedhktfdgdktyqvnvqvgfidegnytsdgnhtltdphyved-kghkylvfe 579
QY	432	ASTGLEN-YOGEQDQYNNWLTNGDDAFNIKSLFRILSNDDIKSRATWANAATGILKNKD 490
Db	580	antgtengyqgeesifnkayyvggtnfrkesqk-lqqsakkrdaelangaigilelnnd 638
QY	491	EKNPKVAELYSPLISAPMVSDIEIRPNVVLGNKYLYLFAATRLNRGSDDDAWNANYAVG 550
Db	639	ytllkvmk---plitsntvtdeieranvfmngkwyfltdsgskmtldg--insn---- 689
QY	551	DNVAMGVVADSLGSKYKPLNDSGWLTASVPANWRTATYSYAVPVACKDDQVLVTSYM 610
Db	690	---diymlygvsnsllgpykpkinktgylvlqmldpndvtfyfshfavpqa-kgnvvvitsym 747
QY	611	TNRNGVAGKGMDSWTWAFSPFLQINPDNTTVLAKMTNQG 649
Db	748	tnrgffedk--katfapslmnikgnktsvvnksilleqg 784
RESULT	9	
ABG21602	ID	ABG21602 standard; Protein; 1095 AA.
XX	AC	ABG21602;
XX	AC	ABG21602;
DT	XX	18-FEB-2002 (first entry)
XX	XX	Novel human diagnostic protein #21593.
DE	XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	XX	Homo sapiens.
OS	XX	
XX	XX	WO200175067-A2.
PN	XX	11-OCT-2001.
PD	XX	
XX	XX	30-MAR-2001; 2001WO-080631.
PF	XX	
PR	PR	31-MAR-2000; 2000US-0540217.
PR	PR	23-AUG-2000; 2000US-0649167.
XX	XX	(HYSE-) HYSEQ INC.
PA	PA	
PI	PI	Drmanac RT, Liu C, Tang YT;
DR	DR	WPI: 2001-639362/73.
XX	XX	N-PSDB; AAS85789.
PT	PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	PT	diagnostics, forensics, gene mapping, identification of mutations
PT	PT	responsible for genetic disorders or other traits and to assess
PT	PT	biodiversity
PS	PS	Claim 20; SEQ ID No 51961; 103pp; English.
XX	XX	
CC	CC	The invention relates to isolated polynucleotide (I) and
CC	CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	CC	and gene mapping.

```
Query Watch      18.4%   Score 759.5; DB 22; Length 1095;
Best Local Similarity 32.3%; Pred. No. 1.5e+38;
Matches 221; Conservative 118; Mismatches 252; Indels 93; Gaps 25;

QY    4  SCKNNAVTLSTAA-----LVFGATTYNASADTNIENNDSSTVQVTTGDND 49
Db    462 asnswwsvtlgtdeapavevpyvgirlvegfilkppmamchrmgvrpavplltqgs 521

QY    50 IAVKSVTLGSGGVSAASDTTIRTSANANGASSAAN---TONSNSQOVASSAITSTSSEA 105
Db    522 -----gegkdsgtphslhtkaqlpshvirhgqvlrrqhaklvgtkaistcgkal 573

QY    106 ASLNTSDSKAAQENTIAKNDDTKAPANESEAKNEPAVNVDSSAAKNDDOQSRRK- 164
Db    574 rtl-----ptak--vfislpnp--ldfkvapslkrksirednrgsrsktv 616

QY    165 NTTAKLNKDABENVVKAGI---DPNSLTDQIKALNMNFMSKAASKG---TQMVTNDFOK 218
Db    617 htegdmmmlkkikvatvltftallagatqafakenngkayketygvshirhdmlq 676

QY    219 IADTLRIQDGRYTVPPFKASEIKMPEAATTKDAQNTTIEPLDVSDSWPQVDRVGQVANW 278
Db    677 lpkq--cqnekyyqvqfsgstiknlsea-----kgldvwdswplqna-dgtvaey 723

QY    279 NGYOLVITAMGIP- NQNNDHIILYLNKYGDNLSHWKNVGPI-----FGYNSTAV---S 328
Db    724 ngynhvrfalagspkdaddtslmyfyqkvgnsdswknagnrvfkdsdkfdandpilkdqt 793

QY    329 QEMSGAVLNSDNSIOLFTRYTRVDTSDNNTNQKTASATLYLTLDNGNVSLAQVRNDYIVF 388
Db    784 qewsgsatfsdgkrlrlyt--dysgkhvyqksittcaqnvsaddtiklnrvdhktif 841

QY    389 EGGGYYYQYDQW--KATNKGADNTAMDRAHVIEDGNRGDRVLVEASTGLEN-YQGEDI 445
Db    842 dggdktyqnvvqqfidegnrytsdghltrdphýved-kghkylvfeantgtengyqgeesl 900

QY    446 YNKLWTGGDDAFNIKSLPRILSNDDIKSRATWANAATGILKLNDKENPKVAELYSPLIS 505
Db    901 fnkayyggntnfrrkesqk-lqqasakrdaelangaigielndnytlkkvmk---plit 956

QY    506 APVMSDEIERPNVVYLGNNKYLYFAATRLNRGSNDDAWMNNAVAVGNVMVGVADSLTG 565
Db    957 snctvdeieranvfkmngkwlyftdsrgskmtldg--insn-----diymlygyvnsltg 1009

QY    566 SYKPLNDSGVWLTAISPANWRRTATSYVAYVPVACKDQOVLVTSYMTNRNGVACKGMDSRW 625
Db    1010 pykpInktglvlgmqldpnvdvtfysfvppqa-kgnnvvitsymtnrgfnfedk--katf 1066

QY    626 ARSFLLQINPDNTVTTLAKMTNQ 649
Db    1067 apsfllminkngktsvkvksilepq 1090
```

```

Db 522 -----gegdsgtphshtkaqlpsphvhrhqqvrrqhsqkivgtkaistgkal 573
Qy 106 ASLNTDSKAAQENTAKNDTQKAAAPANESSEAKNEPANNVNDSSAAKNDQOQSSKK- 164
Db 574 rtl-----pkac---vflsppn--ldfkvapsilkprksirednrgsrqktv 616
Qy 165 NTTAKLNKDAENVVKAGI---DPNSLTDDQIKALKNFNSKAAKSG---TQMTYNDQFK 218
Db 617 htegdmmnlkikvqatvltftallagatqafakennqkayketygvshtrhdmql 676
Qy 219 IADTLIKQDGRYVPFKAISEIKNMPAATTKDAQTNTIEPLDWDWSPVQDVRTGOVANN 278
Db 677 ipkq--qnekyvvpfdgdstikniesa-----kgldwdswpqlna--dgtvaey 723
Qy 279 NGYQLVIAMGIP--NQNDNHIYLLYKNYGDNELSHWKNVGPi-----FGYNSTAV---S 328
Db 724 ngyhvvfalagspkdaddtsiymfykvgdnsidswnknagrvmfkdskfdandpilkdqt 783
Qy 329 QEWSSGSAVLNSDNIOLFTRVDTSDNNTNHQIKASATLYLTDNNGVSLAOVRNDYIVF 388
Db 784 qewsgsatfsgdkirlyft--dysgkhygkslttaqvnvksdtdtkinrvdhktif 841
Qy 389 EGDGYIYQYTDQW--KATNKGADNTAMRDVDAHVIEDGNGDRYLVFEASTGLEN-YOGEQDI 445
Db 842 dgdktyqnvqgfidegnytsghdhtirdphived-kghkyivfeantgtengyqgeesl 900
Qy 446 YNWLNYGGDDAFNIKSLFRILSDNDIKSRATWANAAIGILKLNKDEKNPKVAELYSPLIS 505
Db 901 fnkayvggntnfrkesqk-lqgsakkrdaelangelgielndytlkkvmk---pllt 956
Qy 506 APWVSDTERPNVVKLGNYKLYLFAATRLNRSNDNDAMNANVAVGDNVAMGYVADSLTG 565
Db 957 sntvtdeleranvfmngkwyiftdsrsgskmtidg--lnsn-----dlymgyvsnsltg 1009
Qy 566 SYKPLNDSGVVLTSVPAWNRATATYSYVAVPVAGKDDQVLTYSYMTNNGVAGKGMDSW 625
Db 1010 pykplnkglvqlmgldpndvftysfhavpqa-kgnnvitsymtngffedk--katf 1066
Qy 626 APSFLQLQNPNTTYTLAKMTNQG 649
Db 1067 apsfllmnikgnktsvkvnsilepg 1090

RESULT 11
ID ABG04307
XX ABG04307 standard; Protein; 527 AA.
XX AC ABG04307;
XX XT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #4298.
XX KW Human; chromosome mapping; gene mapping; genetic therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.

```

```

DR N-PSDB; AAS68494.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 34666; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC imaging of sites expressing (II). (I) and (II) are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 527 AA;

```

```

Query Match 18.4%; Score 758.5; DB 22; Length 527;
Best Local Similarity 37.6%; Pred. No. 6.5e-39;
Matches 194; Conservative 90; Mismatches 181; Indels 51; Gaps 19;
Qy 155 KNDDQOQSSKK--NTTAKLNKDAENVVKAGI---DPNSLTDDQIKALKNFNSKAAKSG- 208
Db 37 rednrgsrqktvhtegdmmnlkikvqatvltftallagatqafakennqkaykety 96
Qy 209 --TQMTYNDQFKIADTLIKODGRYVPFKAISEIKNMPAATTKDAQTNTIEPLDWDWSP 266
Db 97 gvshtrhdmqlipkq--qnekyvvpfdgdstikniesa-----kgldwdswp 144
Qy 267 VQDVRTGOVANNVYQVLIAMGIP--NQNDNHIYLLYKNYGDNELSHWKNVGPi----- 319
Db 145 lqna-dgtvaeyngyhvvfalagspkdaddtsiymfykvgdnsidswnknagrvmfkdsk 203
Qy 320 FGYNSTAV---SOEWSGSAVLNSDNIOLFTRVDTSDNNTNHQIKASATLYLTDNNGV 376
Db 204 fdandpilkdqtqewsgsatfsgdkirlyft--dysgkhygkslttaqvnvksdtdt 261
Qy 377 SLAOVRNDYIVFEGDGYIYQYTDQW--KATNKGADNTAMRDVDAHVIEDGNGDRYLVFEAST 434
Db 262 kingvedhktifdgdgktyqnvqgfidegnytsghdhtirdphived-kghkyivfeant 320
Qy 435 GLEN-YOGEQDIYNWLNWNYGGDDAFNIKSLFRILSDNDIKSRATWANAAIGILKLNKDEKN 493
Db 321 gtengyqgeeslfnkayvggntnfrkesqk-lqgsakkrdaelangelgielndytl 379
Qy 494 PKVAELYSPLISAPWVSDTERPNVVKLGNYKLYLFAATRLNRSNDNDAMNANVAVGDNV 553
Db 380 kkvmk---plitsntvtdeleranvfmngkwyiftdsrsgskmtidg--lnsn----- 429
Qy 554 AMGVTVADSLTGSYKPLNDSGVVLTSVPAWNRATATYSYVAVPVAGKDDQVLTYSYMTN 613
Db 430 ymlgyvsnsltgpykplnkglvqlmgldpndvftysfhavpqa-kgnnvitsymtng 488
Qy 614 NGVAGKGMDSWAPSFLQLQNPNTTYTLAKMTNQG 649
Db 489 gffedk--katfapsflmnikgnktsvkvnsilepg 522

```


RESULT 12

ABG23205

ID ABG23205 standard; Protein; 789 AA.

XX

AC ABG23205;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #23196.

XX

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

KW

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

FR

XX

23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

Drmanac RT, Liu C, Tang YT;

PI

XX

WPI; 2001-639362/73.

DR

XX

N-PSDB; AAS87392.

DR

XX

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

PT

XX

Claim 20; SEQ ID No 53564; 103pp; English.

PS

XX

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

CC

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

CC

XX

Sequence 789 AA;

XX

Query Match 18.48; Score 758.5; DB 22; Length 789;

Best Local Similarity 37.68; Pred. No. 1.1e-38;

Matches 194; Conservative 90; Mismatches 181; Indels 51; Gaps 19

QY 155 KNDQOQSKK--NTAKLNKDAENWVKAGI---DPNSLTDDIKALKNMFSAKSG- 208

DB 299 rednrgsqkvhtegdmnmikivkqatvltfttallagatqafakennqykety 358

Qy	377	SLAQRNDYIVFEGDGYYYQTDQW--KATNKGADNIAMRDAHVIEDGNGDRYLVFEAST	434
Db	228	kingvedhktifdgdgkytynvgvqfiddegnytsgdnhtrdrphyeved-kghkylvfeant	286
Qy	435	GLEN-YGEGDOIYNWLNIGYGGDDAFNKSUFRLISNDNDIKSRATWNAATGIILKLNKDEKN	493
Db	287	gtengygeeslnfkyaygg9gtffkfesqk-lqgsakkrdaelangeliielnndytl	345
Qy	494	PKVAELYSPLISAPWVSDETERPNVVKLGKYLFAATRLNGSNDDMANNAVAVGDNV	553
Db	346	kkvmk---plitsntvtdeleranvfkmngkwyflfcdsrgskmtldg--insn-----di	395
Qy	554	AMVGYVADSLTSGKPLNDSDGVVLTVASVPANNRTATYSYAVVPVAGKDQDVLTYSMTNR	613
Db	396	ymlgyvshsltgykplnktglvlgmgldpndvftfysnhfvpgq-kgnnvvitsymtnr	454
Qy	614	NGVAGKGMDSWAPSELLQINPDNTTTLAKMTNQG	649
Db	455	gfedk--kattapsflmknkgkntsvvknslleaq	488

RESULT 15

RESOL 13
ABG21832

ID ABG21832

XX
SECRET SCANDAL, PROLEIN

AC ABG21832;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human dia

XX
11:50 - 1

KW Human; chromosome mapping; 9

XX
KW
rood supplement; medical ima

Homo sapiens

XX
nono sapientis.

PN WO200175067-A2.

XX
XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

[illegible]

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
yy

XX-
XX-
PA (HVSE-1) HVSCC YNC

PA (HYSE-) HYSEQ INC.
XX

Drmanac RT Lin C Ena 3 v6-

LIU C, Tang YT;
XX

DR WPI: 2001-639362/73

DR N-PSDB; AAS86019.

XX
XX

PT New isolated polynucleotide and encoded polynucleotide

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits and to assess

biodiversity -

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

Claim 20; SEQ ID No 52191; 103pp; English.

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX Sequence 546 AA;

Query Match 18.3%; Score 756; DB 22; Length 546;
Best Local Similarity 36.9%; Pred. No. 9.7e-39;
Matches 105.2

```

141 QY KNEPAVVNVNDSSAAKNDDOQSSKK--NTTAKLNKDAENVVKAGI---DPNSLTDQIQKAL 196
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
43 Db kvapsilpkrsirednrgsrqkvthtegdmnnnikivkqatvltfttallagatqaf 102
    QY NKWNFSKAAKSG---TQMTYNDPQKADTTLIKQDGRYTVPFPEKASETKNNPAATTKDAQOT 253
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
103 Db akennqkayketygvshtrndmlqpkq--gqnekypvqpfdqgstikniesa----- 153
    QY 254 NTTEPLDVNDSPVQDVRTGQVANNWGYQIVIAMGIP--NONDNHIIYLLXNKYKGDNELSH 312
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
154 Db ---kgldvdswpqlna--dgtvaeyngyhvfalagspkaddtslymfqkvqvgdnsids 209
    QY 313 WKNVGP1-----FGVNSTAV---SOEWSSAVLNSDNSIQLPYTRVDTSDNNTNHQKIA 363
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
210 Db wknagrvfkdskdfaandpilkdqtdgswsaftsdgkirlft--dysgkhgygkslt 267
    QY 364 SATLYLTDNNGNVSLAQVRNDYIVFEGDGYYYOTYDQW--KATNKGADNIAMRDARHVED 421
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
268 Db taqvnsaddtliknrvedhktifdgdgkyqnvqgfidegnytsghnhtlrdphived 327
    QY 422 GNGDRLYFEASTGLFN--YQGEDOIYNWLNYGDDDAFNKSLPRILNSDDIKSRATWANA 480
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
328 Db -kgkhylvfeantgtengyvggeesifnkayygggtnfrrkesqk-lqgsakkrdaelang 385
    QY 481 AIGILKLNKDEKNPVAELYSPLISAPWSDYEIERPNVVKGLNKKYVLAATRLNRSNDD 540
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
386 Db aligielndndytlckvmk---pltsntvtdeieranvfkmngkwyflfcdsrgskmtidg 442
    QY 541 AMMNANYAVGDNVAMVGIVYADSLTSGYKPLNDSGVVLTASVPANRRTAYSYVAVPVAK 600
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
443 Db --insn-----diymlygvsnsltpgypkplnktgvlqmgldpndvtfcysfravpqa-k 494
    QY 601 DDOVLVTSYMTNRRNGVAGKMGDSTWAPSFLLQINPDNTTTLAKMTNQG 649
    Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
495 Db gnnvitsymtnrgffedk--katfapsflmnikgnktsvvknsilepqa 541

```

Search completed: September 26, 2002, 18:26:15
Job time: 154 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2002, 22:12:59 ; Search time 27.35 seconds
(without alignments)
704.636 Million cell updates/sec

Title: US-09-995-587A-1
Perfect score: 4129
Sequence: .1 MYSGKNWVTLSTALVF.....GMGTLLSFWGLAEINKRRFN 789

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	606.5	14.7	455	2	US-08-870-827-3
2	281	6.8	415	1	US-08-381-936-2
3	281	6.8	415	3	US-08-943-374-2
4	279	6.8	578	4	US-09-503-172A-2
5	266	6.4	543	1	US-08-362-232-2
6	266	6.4	543	1	US-08-814-196-2
7	254.5	6.2	1164	4	US-08-923-992A-10
8	246.5	6.0	1128	4	US-08-923-992A-6
9	242	5.9	984	1	US-08-242-932-2
10	242	5.9	984	1	US-08-714-481-2
11	242	5.9	984	5	PCT-US95-06111-2
12	241	5.8	1184	4	US-08-923-992A-2
13	240.5	5.8	1098	4	US-08-923-992A-8
14	238	5.8	1104	4	US-08-923-992A-4
15	220	5.3	826	1	US-07-638-431-2
16	220	5.3	826	5	PCT-US92-00018-2
17	213	5.2	1565	6	5352450-2
18	194	4.7	1912	1	US-08-409-995-4
19	194	4.7	1912	3	US-08-685-467-4
20	194	4.7	2411	4	US-09-268-347-36
21	192	4.7	2353	4	US-09-377-155-33
22	192	4.7	2353	4	US-08-913-942-4
23	192	4.7	2353	4	US-09-669-974-33
24	191.5	4.6	1561	3	US-08-894-017-23
25	189	4.6	2354	4	US-09-268-347-47
26	184	4.5	1430	3	US-09-008-172-2
27	184	4.5	1430	4	US-09-210-361-6

28	183.5	4.4	2314	4	US-09-268-347-49	Sequence 49, Appl
29	183	4.4	1566	2	US-08-687-956A-23	Sequence 23, Appl
30	182	4.4	1296	3	US-08-470-260-3	Sequence 3, Appl
31	182	4.4	1296	3	US-08-471-491-3	Sequence 3, Appl
32	182	4.4	1296	4	US-08-466-662-3	Sequence 3, Appl
33	181.5	4.4	1094	4	US-09-268-347-32	Sequence 32, Appl
34	180.5	4.4	1702	4	US-08-296-791-5	Sequence 5, Appl
35	180.5	4.4	1702	5	PCT-US95-10661A-5	Sequence 5, Appl
36	179.5	4.3	1287	1	US-08-200-232-2	Sequence 2, Appl
37	179.5	4.3	1287	5	PCT-US95-02219-2	Sequence 2, Appl
38	179.5	4.3	1287	5	PCT-US95-02219A-2	Sequence 10, Appl
39	178	4.3	2123	4	US-08-968-685A-10	Sequence 2, Appl
40	176.5	4.3	1577	2	US-08-793-824-2	Sequence 2, Appl
41	175.5	4.3	933	3	US-08-293-728-2	Sequence 2, Appl
42	175.5	4.3	933	3	US-09-421-868-2	Sequence 2, Appl
43	170	4.1	1183	2	US-08-447-031A-2	Sequence 2, Appl
44	168	4.1	1545	4	US-08-296-791-4	Sequence 4, Appl
45	168	4.1	1545	5	PCT-US95-10661A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-870-827-3
; Sequence 3, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; * TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 455 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-870-827-3

Query Match 14.7%; Score 606.5; DB 2; Length 455;
Best Local Similarity 36.2%; Pred. No. 4.2e-36;
Matches 156; Conservative 75; Mismatches 147; Indels 53; Gaps 17;
OY 229 RYVPPFKASEIKNMPAAATTKDAQTNIPLVDVWDSVPQDVRGTGVANNNGVQLVIAMM 288
:: ||| ||| ||| : | : |||||::| : ||::| :| :

Db 31 QFKVPQFNASAKIDSADKSGNLD-LDWDWSPLQNA-DGTAANYHGTHVSALA 88
 QY 289 GIP-NONDNHLYLLYKNGDNLHKNVGPPIFYNSTAV-----SQEWSSGSAVLN 338
 Db 89 GPKNSDDTPLHLFYQKVGQTSIDSKNAGRFEEDMDKFPNDPYLYKYQTEWSSGSAVLN 148
 QY 339 SDNSIQLFYTRVDTSDN-----NTHQKIASATLYLTD-NNGNVSLAQRNDYIVFE-G 390
 Db 149 KDGVRLEFY--DYSNPEDGGTGAGNQIISTAQVNLSPQDAATLKVDGVSDBKSVFDDG 206
 QY 391 DGYVYQYDQWKATNK--CADNIAMDRAHVEDGNGDRYLVEASTG--LENYOGEDQIYN 447
 Db 207 DGTVYQIQOQFIDEGKWISGDNHTLRDPHYVED-KGHKYLVEANTGTTDGVQGGQSQNN 265
 QY 448 WLNYGGDDAFNIKSLFRILSNDNDIKSRATWANAAGILKLNKDKNPVVAELYSPLISAP 507
 Db 266 KAYYGGSDVFFQNEKNKLLQSPK-QKIASLANGALGIIVELADDY---TVKSMKPLVASN 321
 QY 508 MYSDETERPNVYKLNKYLFLFAATRLNRGSNDAMNANYAVGD-NVAMVGYVADSLTGS 566
 Db 322 TVADEVERANIFKMNKWLFTDSRGSMTSD-----GINDKDYVYMLPGGDSLNGP 373
 QY 567 YKPLNDGSGVLTASYPANWRTATSYAVPVAGKDDQVLVTSYNTNRN----- 614
 Db 374 HNPINETGLVNMNLDPADLTHYSHCGIP-HPGNNVYVLTSTYNTNRGFPYPERHSHLRDK 432
 QY 615 -GVACKGMDST 624
 Db 433 LGVNIKGS DTS 443

RESULT 2

US-08-936-2
 ; Sequence 2, Application US/08381936
 ; Patent No. 5792923

GENERAL INFORMATION:

APPLICANT: ROBER, Manuela
 APPLICANT: GEIER, Gebhardt
 APPLICANT: GEIDER, Klaus
 APPLICANT: WILLMITZER, Lothar
 TITLE OF INVENTION: DNA sequences which lead to the
 TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
 TITLE OF INVENTION: these sequences as well as a process for preparing
 TITLE OF INVENTION: transgenic plants.
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
 STREET: 1180 Avenue of the Americas
 CITY: New York
 STATE: NY
 COUNTRY: US
 ZIP: 10036-8403
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/381.936
 FILING DATE: 09-FEB-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 93 02110
 FILING DATE: 09-AUG-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: DE P42270618
 FILING DATE: 08-DEC-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Meilman, Edward A.
 REGISTRATION NUMBER: 24,735
 REFERENCE/DOCKET NUMBER: P/951-108
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888
 TELEX: 236925
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 415 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-381-936-2

Query Match

Best Local Similarity 6.88; Score 281; DB 1; Length 415;
 Matches 109; Conservative 73; Mismatches 137; Indels 106; Gaps 22;

QY 257 EPLDVWSPVQDVRTGQVANWNGYQLVIAAMGIPN-----QNDNHIYLL-----YNKYG 306
 Db 40 EEVFINDTMPLRDF-DGEIISVNGWCIIFTLTADRTNDNPQFODENGNDYDITRDWEDRHG 98
 QY 307 DNELSHW-KNVGP--IFGYNSTA-----VSQEWSSGSAV-LNSDNSIQLFYTRVDTSDNNT 357
 Db 99 RARICWYSTRGDKWIFGGRVMAEGVAPTTRWAGTPILLNDRGDIDLYITCV-----T 152
 QY 358 NHOKIASATLYLTDNNGVNSLAQRNDYIVFEGDGYVYQYDQ---WKATNKGADNIAMR 414
 Db 153 PGATIAKVRGKIVTSQSVSLEGFQOVTSLFSADGTIYQTEBONAFWN-----FR 202
 QY 415 D-AHVIEDGNGDRYLVEASTGLENYQGBDIY-----NWLNYGGDDAFNIKSLFR 464
 Db 203 DFPSPFIDRNDGKLYMLFEGN--VAGPRGSHEITQAEQNVPPGYEDVG----- 249
 QY 465 ILSNDDIKSRATWANAAGILKLNKDKNPVVAELYSPLISAPVSYDEIERNPVYKLNK 524
 Db 250 -----AKYQAGCVG-LAVAKDLGSGE-WQLPLPLITAVGVNDQTERPHEVFDQCK 297
 QY 525 YVLFATRLNRGSNDAMNANYAVGDVNY-----AMVGYVADSLTGSYKPLNDSGVLTAS 580
 Db 298 YLFT-----ISHKYTFADNLGTGPDGVYGVPSDKLTGPTYPNMSSGLVL--G 342
 QY 581 VPANWRTATSYAVPVAGKDDQVLVTSYMTNRNGVAGKGMV-----STWAPSFLQINPD 636
 Db 343 NPSSQPFQYSHYVMP-----NGLVTSFI---DSVPMKGDYRIGGTGTEAPTIVKLLKGD 393
 QY 637 NTTV 641
 Db 394 RSFIV 398

RESULT 3

US-08-943-374-2
 ; Sequence 2, Application US/08943374
 ; Patent No. 6028249

GENERAL INFORMATION:

APPLICANT: ROBER, Manuela
 APPLICANT: GEIER, Gebhardt
 APPLICANT: GEIDER, Klaus
 APPLICANT: WILLMITZER, Lothar
 TITLE OF INVENTION: DNA sequences which lead to the
 TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
 TITLE OF INVENTION: these sequences as well as a process for preparing
 TITLE OF INVENTION: transgenic plants.
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
 STREET: 1180 Avenue of the Americas
 CITY: New York
 STATE: NY
 COUNTRY: US
 ZIP: 10036-8403
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICANT: TONOKURA, Takashi
;   APPLICANT: SAKANO, Yoshiyuki
;   TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE
;   FILE REFERENCE: 10749-0001-0
;   CURRENT APPLICATION NUMBER: US/09/503,172A
;   CURRENT FILING DATE: 2000-02-14
;   PRIOR APPLICATION NUMBER: JP 160416/1999
;   PRIOR FILING DATE: 1999-06-08
;   NUMBER OF SEQ ID NOS: 9
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 2
;   LENGTH: 578
;   TYPE: PRT
;   ORGANISM: Arthrobacter sp.
; US-09-503-172A-2

```

```

Query Match      6.8%; Score 281; DB 3; Length 415;
Best Local Similarity 25.6%; Pred. No. 1.4e-12;
Matches 109; Conservative 73; Mismatches 137; Indels 106; Gaps 22;

QY 257 EPLDWDSPVQDVRTGQVAMNGYQYLVIAAMGIPN-----QNDNHVLL-----YKYG 306
DB 40 EEVFIWMTPLRDF-DGEIISVNGWCIIFTLTADRNTDNPQDQNGNDYDTRDWDHRG 98

QY 307 DNLSHW-KNVGP--IPGYNSTA-----VSQWGSQAV-LNSDNSTQLYTRVYTSNNNT 357
DB 99 RARICWYSTGDKWIFGGRVMAEGVAPTTREWAGTPILLNDRGIDLYITCV-----T 152

QY 358 NHQKIASATLYLTDNNGNYSVLAQVRNDYIVFEGDGYYYQYDQ---WKATNKGADNIAMR 414
DB 153 PGATIAKVRGKIVTSDQSVSLEGFOQVTSLSFADGTIYQTEQNAFN-----FR 202

QY 415 D-AHVIEDGNGRYLVFEASTGLENYOGEDQIY-----NWLYNGDDAFNIKSLFR 464
DB 203 DQSPFIDRNDGKLYMLFECN--VAGPRGSHETQAEQNVPPGYEDVGG----- 249

QY 465 ILSNDDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPMVSDIERPNVVKLGK 524
DB 250 -----AKYQAGCVG-LAVAKDLGSE-WQILPLITAVGVNDQTERPHFVQDGK 297

QY 525 YLFAATRLNRGSDNDAMNANYAVGNV-----AMGVYVADSLTGSYKPLNDSGVILTAS 580
DB 298 VYLF-----ISHKYTFADNLTPGDGVYGVSKLTPGYPNNSGLVL--G 342

QY 581 VPANWRATYTYAYVAVPAGKDDQVLTYSYMTNRNGVAGKGM-----STWAPSELLQINPD 636
DB 343 NPSSQEPQTYSHYVMP-----NGLVTSFI---DSVPWKGKDIRIGTEAPTVKILLKGD 393

QY 637 NTTFV 641
DB 394 RSFTV 398

```

```

RESULT 4
US-09-503-172A-2
; Sequence 2, Application US/09503172A
; Patent No. 6284510
; GENERAL INFORMATION:
; APPLICANT: ITO, Tetsuya
; APPLICANT: FUJITA, Koki
; APPLICANT: HARA, Kozo

```

```

; APPLICANT: TONOKURA, Takashi
; APPLICANT: SAKANO, Yoshiyuki
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE
; FILE REFERENCE: 10749-0001-0
; CURRENT APPLICATION NUMBER: US/09/503,172A
; CURRENT FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 160416/1999
; PRIOR FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Arthrobacter sp.
; US-09-503-172A-2

```

```

Query Match      6.8%; Score 279; DB 4; Length 578;
Best Local Similarity 23.4%; Pred. No. 3.3e-12;
Matches 139; Conservative 83; Mismatches 207; Indels 164; Gaps 28;

QY 131 APANESSEAKN--EPAVNVNDSSAAKNDQSSKNTTAKLNKDAENVVKKAGIDPNSL 188
DB 31 AAPATAQSGLDGPEPTIHTQAYAPEDD-----FTAK----- 63

QY 189 TDDQIKALKNMFSKAASGTOYTNDFOKIADTLIKQDGRYVTPFFKASEIKNMPAATT 248
DB 64 -----WTRADARQLQMSDP-----TAP-----SRENSWPASVT 92

QY 249 -----KDAQNTIEPLDWDSPVQDVRTGQVAMNGYQYLVIAAMGIPNQ--NDNHY-- 299
DB 93 MPTVPQDPEPDMSEQVWWDWTLTDEDANQ--SVNGWEIIFSLVADRNLGDFDRHVF 151

QY 300 --LLYKYGDNLSHWK-----VGPFGYNSTAVSQWGSQAVLNSDSNI 343
DB 152 IGYFRPAGVPAAPERPENGWGTGGLVFKGVTGQIFEDQSFSHOTQWSSARVSKNGEI 211

QY 344 QLFYTRV-----DTSNNTNHQKIASATLYLTDNNGNYSVLAQVRNDYIVFEGDGYYYQ 397
DB 212 KLFFTDVAFYRNSDGTNIKPYDPRIALSVKVKANKKGVTLTGFKNKTDLQADGTYQT 271

QY 398 YDQNKATNKGADN--IAMRDAHVIED--GNGDRYLVFEASTGLENYOGEDQIYNWLN 452
DB 272 -----GAONEFFNRPDPTFEDPAHPGETFMVEGNSAM---QRETATCNEADLGY 319

QY 453 --GDPAFIKSLRILNSDDIK--SRATWANAAGILKLNKDEKNPKVA--ELYSPLISAP 507
DB 320 RQGDPAE-----TVDDVNASGATYQIGNVGLAK-----AKNKOLTEWEEFLPPLSAN 367

QY 508 MVSDEIERPNVVKLGKYYLYLFAATRLNRGSDNDAMNANYAVG--DNVAMGVYVADSLT 565
DB 368 CVTDQTERPQIYFKDGKSYLF--TISHRGT-----FAAGLDGPEGYGVFGDGIRS 416

QY 566 SYKPLN--DSGVVLTASVPANW-----RTATYSYAVVPAGKDDQVLTYSY 609
DB 417 DYQPLNGGSGLAGNPTNLNFGGQPFADFNQHPGHFOAYSHYVMP--GGLVQSFIDTI 474

QY 610 MTRNRGAVKGMDSWAPSFLLQINPDNTTTLA---KMTNOGDWIDWDSSEN 659
DB 475 GTHDDEFVYK-----GTILAPTVKMDIGVGDPTKTAVDYSYSGSEGLGWADIPAN 523

```

```

RESULT 5
US-08-362-232-2
; Sequence 2, Application US/08362232
; Patent No. 5641667
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Fructosyltransferase Enzyme, Method
; TITLE OF INVENTION: For its Production and DNA Encoding the Enzyme.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron

```

```

; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,232
; FILING DATE: 22-December-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-362-232-2

```

```

Query Match          6.4%; Score 266; DB 1; Length 543;
Best Local Similarity 26.9%; Pred. No. 2.6e-11;
Matches 141; Conservative 63; Mismatches 171; Indels 150; Gaps 35;

QY 209 TQMTY---NDPQ---KIADPL-IKQGRYTVPFKASEIKNMPAATK---DAQTNTIEP 258
Db 33 TQAYDPQSDFTARWTRADALQIKAHSDATV-----AAGQNSLPAQLTMPNIPADFVPINP 88
QY 259 -LDVWSWPQVDVYRTGVANNGYQVLVIAMMGIPNQ-----NDNHIY-----LLYNYKGDNE 309
Db 89 DVWVMDTWTLIDKHADQF-SYNGWEVIFCLTADPNAGYGFDDRHVHARIGFFYRRAGIPA 147
QY 310 LSH-----WKNVPIFGYNSTAV-----SOEWSSGSAVLNS--DNSIQLFYTRV--- 350
Db 148 SRPVGNGWTYGGHLFPDGSAGQVYAGQTYTNOAEMSGSSRLMQIHGNTVSVFYTDVAFN 207
QY 351 -DTSNN-TNHOKIASATLYLTDNNGVNSLAQVRNDY--IVFEG-----DGYYYQ 396
Db 208 RDANANNITPPQAIIQT-----LGRHADFNHVMFTGTFTHTAHTPLLPDGVLYQ 256
QY 397 TYDQWKATNKGADN--IAMRDAHVIEDGN--GDYLYLFEAST-----GLENYQGEDQIYNW 448
Db 257 -----NGAQNEFFNFRDPFTEDPKHPCGVNVMVEGNTAGORGVANCTEAD----- 302
QY 449 LNYGGDAFNKSLFRILSNDDIKSRATWANAAGILKLNKDEKNPKVAELYSPLISAPM 508
Db 303 LGFRPNP-NAETLOEVLN-----SGAYIQANIG-LAIATDSTLSK-WKFLSPLISANC 354
QY 509 VSEITERPNVYKLGKYYLFAATRLNRGSDDDAWNANYAVG-----DNVAMVGVADSLT 564
Db 355 VNDQTERPQVYLHNGKYYFIETISH-----RTFAAGVDGPGV--YGEVGDGIR 401
QY 565 GSYKPLNDGSGVVLATASVPANRHTA-----TYSYAVPVAGKDDQVLVT 607
Db 402 SDFQPMN-YGSLTNGNPTDLNLTAAAGTDFDPSDQNPRAFQSYSHYVMPGG-----LVE 454
QY 608 SY---MTNRNGVAGKGMDSWAPSFLLQINPDNTTTVLAKMNOG 649
Db 455 SFIDTVENRRG-----GTLAPTIVRYRI-AQNASAVDLRYNGG 491

```

```

RESULT 6
US-08-814-196-2
; Sequence 2, Application US/08814196
; Patent No. 5731173
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Fructosyltransferase Enzyme, Method
; TITLE OF INVENTION: For its Production and DNA Encoding the Enzyme.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,196
; FILING DATE: 10-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,232
; FILING DATE: 22-December-1994
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-814-196-2

```

```

Query Match          6.4%; Score 266; DB 1; Length 543;
Best Local Similarity 26.9%; Pred. No. 2.6e-11;
Matches 141; Conservative 63; Mismatches 171; Indels 150; Gaps 35;

QY 209 TQMTY---NDPQ---KIADPL-IKQGRYTVPFKASEIKNMPAATK---DAQTNTIEP 258
Db 33 TQAYDPQSDFTARWTRADALQIKAHSDATV-----AAGQNSLPAQLTMPNIPADFVPINP 88
QY 259 -LDVWSWPQVDVYRTGVANNGYQVLVIAMMGIPNQ-----NDNHIY-----LLYNYKGDNE 309
Db 89 DVWVMDTWTLIDKHADQF-SYNGWEVIFCLTADPNAGYGFDDRHVHARIGFFYRRAGIPA 147
QY 310 LSH-----WKNVPIFGYNSTAV-----SOEWSSGSAVLNS--DNSIQLFYTRV--- 350
Db 148 SRPVGNGWTYGGHLFPDGSAGQVYAGQTYTNOAEMSGSSRLMQIHGNTVSVFYTDVAFN 207
QY 351 -DTSNN-TNHOKIASATLYLTDNNGVNSLAQVRNDY--IVFEG-----DGYYYQ 396
Db 208 RDANANNITPPQAIIQT-----LGRHADFNHVMFTGTFTHTAHTPLLPDGVLYQ 256
QY 397 TYDQWKATNKGADN--IAMRDAHVIEDGN--GDYLYLFEAST-----GLENYQGEDQIYNW 448
Db 257 -----NGAQNEFFNFRDPFTEDPKHPCGVNVMVEGNTAGORGVANCTEAD----- 302

```


QY 449 LNYGDDAPNKSIFRLSNDIDKSRATWANRAIGILKLNKDEKNPKVAELYSPLISAPM 508
 Db 303 LGRPNP NAETLQVLD-----SGAYOKANIG-LAIATDSTLSK-WKFLSPLISANC 354
 QY 509 VSEIERPNVYLGKYYLFAATRLNRGSDDAWNNYAVG-----DNVAMVGYVADSLT 564
 Db 355 VNDQTERPOVYLHNGKYYIFTISH-----RTTFAAGVGDGVR--YGFVGDGIR 401
 QY 565 GSYKPLNDSSVVLTSVPAWRTA-----TYSYVAVPVAGKDDQVLVT 607
 Db 402 SDFQPMN-YGSLTGMGNPTDLNTAAGTDFDPSQDNPRAFOSYSHYVMPGG-----LVE 454
 QY 608 SY---MTNRNGVAGKGMDSWAPSLLOINPDNTTTLVAKTNOG 649
 Db 455 SPIDTVENRRG-----GTLAPTVRVRI-AQNASAVDLRYGNGG 491

RESULT 7

US-08-923-992A-10
 ; Sequence 10, Application US/08923992A
 ; Patent No. 6280738
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
 ; TITLE OF INVENTION: Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/923,992A
 FILING DATE: 05-SEP-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/024,707
 FILING DATE: 06-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893
 REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1164 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-923-992A-10

Query Match 6.2%; Score 254.5; DB 4; Length 1164;
 Best Local Similarity 20.2%; Pred. No. 5.6e-10;
 Matches 193; Conservative 135; Mismatches 341; Indels 287; Gaps 41;
 QY 10 VVTLSTAAVFGA-----TTVNAS-----ADTNINNDST-----VQV 43
 Db 21 VASVAVASLFPMSVAHASELVKHSVKRTTEVAAPKPPSMAQTDOGNSSSELETTKIEI 80
 QY 44 TTGDNDIAVKSVTILGSGOVSAASDITRTISA-----NANSASSANTONSQVASSAAIT 99

Db 81 PTTDIKAVEPLEKTAGETS-ATDTGKREKQLOQWNNLNKNDVHNTILSHQKNKNEFTKI 139
 QY 100 SSTSSAASL-----NNT-----DSKAAQENTNTAKNDDTKAAPANESS 138
 Db 140 DETNDSADALLENOFNETNRLHLHIKOEVEVEKDKAKQO--KTLKQSDT-KVDLSNIDK 196
 QY 139 EAKNEPAV-----NVNDSSAAKNDD-----QSSKK-----N 165
 Db 197 ELNHQKSOVETMARQOLITNEDKDSMLKKIEDIRKQQAQADKEDAEVKVREELGKLFTS 256
 QY 166 TTAKLNDKDAENVYKK-----AGID-----PNSLTDQDQIKALNMNFSAKASGTOYNT 214
 Db 257 TKAGLQDEIOEHVKETTSEENTQKVDEHYPNSLQNLAKSLBELDKATTNEQATQV-N 315
 QY 215 DF-----OKIAD--TLIKQDGRYTVPPFFKA-----SEIKNMPAATTKDAQTNTIEPL 259
 Db 316 QFLENAQKLKEIQLIKET--NVKLYKAMSESLQVEKOLKHNQANLEDLVAKSKEIV 372
 QY 260 DVWDS-----WPVQDVRTGQVANWNG 280
 Db 373 REYEGKLNOSKNLPKOLLEEAHSLKQVVEDFRKKEKTSQVTPKRVKRDLAENNN 432
 QY 281 YQIVIAMMGINP-----QNDNHIIYLLYKNKYGDNELSHKNGVPIFGYNSTAVSQWMSG 336
 Db 433 QOKIELTVSPENITVYEGEDLAKFTLTAKSDSKTTLDFSDL--LTKN-PSVSDRISTNYK 489
 QY 337 LNSDNSIQLFYTRVDTSDNNTNHQKIASATLYLTDNNGNVSLAQRNDYIVFEGDGYVQ 396
 Db 490 TWTDN-----HKIAETITIKNLNESQTVTLKAKDSDGNV--VQRTFTTV-----533
 QY 397 TYDQWKATNKGADNIAMRDHVI-----DNGDRYLVFVFEASTGLNLYQGEDQIYN 447
 Db 534 ---OKKEEKQVPTPEQKDSKTEEKVPOEPKSNKNOQLKLSAQOQLEKLE-----583
 QY 448 WLNYGDDAFNKSIFRLSNDIDKS-----RATWAN-----AAGTLKLNKDEKN 493
 Db 584 -----KAIKELMEQPEIPSPNEPYGIQKSINESQKEPIQEAITSEFKKIIGDSS 631
 QY 494 PKVAELYSPLISAPMVSDEIERP-----NVVKGNGKYYLFAATRLNRGSDDAWNNY 547
 Db 632 KYYTEHYFNKYKSDFMVYOLHAQEMLTRKVVQVYINKYIPDIAEIKKIFESDMKRTEDNY 691
 QY 548 AVGNVAMGVVADSLTGSYKPL-----NDSGVLTASVPANWRRTATSYVAVPVAGKDDQ 603
 Db 692 GSELDALGKYFEKYFLTPENKIKQIVDDLDKQVQDQPA-----PIPENSEMDO 741
 QY 604 -----VLVTSYMTNRNGVAGKGMDSWAPSLLOINPDNTTTLVAKMTNQGDWIWDDSS 657
 Db 742 AKEKAKTAVSKYMS-----KVLDDGVHQ-----HLQKKNNTRIV-----DLFKELEA 782
 QY 658 ENLDMIGDLSAALPGERPKVPDWDLIG-----YGLK-----PHDPATNDPPTPTT 704
 Db 783 IKQQTIFIDIDNAKTEVEIDNLVH-DAFSKNATVAFQKGLTNTPTETPTKIPELPOA 841
 QY 705 PETPETNPPTKPTPENPGTPTPTNTPTPEIPLTPETPKQPTQTNRLPQTGN 760
 Db 842 PDTPOADPTPHVPESPKAPEAPRPVPSPTPEAPHVPQSPKAPAP---RVPESPN 894

RESULT 8

US-08-923-992A-6
 ; Sequence 6, Application US/08923992A
 ; Patent No. 6280738
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
 ; TITLE OF INVENTION: Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington

QY	6	KWAVVTJSTAALVFGATTYNASADTNIENNDSST---	VOVTTGDNDIAKVSVTJGS	59
Db	6	KDVSYKTTVEAAKY-----PSMAQTDGNGNSSSELETRRHEIPTTIDIKKAVEPEKTA	60	
QY	60	GOVSAASDTTIRTSA-----NANASSAANYTQNSNOVASSAAITSTSSAASL---	108	
Db	61	GETS-ATDTGKREKQLOQWKNLKNLDVNTILSHOEKNEFKTIDETNDSALLLENOF	110	
QY	109	NNT-----DSKAAQENTNTAKNDOTO-----KAAPANESSAK-	141	
Db	120	NETNRLHLHIKQHEEVERDKKAKQO---KTLQSDTKVDLSNDKELHOKSVEPKMAEPKG	177	
QY	142	--NEPVAVNDSSAAKND---QOSSKKNYTKALNKDAENVYK-----KAGID	184	
Db	178	ITNED---KDSMLKIEDIRKQAQADKK-----EDAEVKVREELKGLFSSTKAGLD	226	
QY	185	-----PNSLTDDQIKALNKNFNSKAAKSGTQWTYNDP-----	216	
Db	227	QEIHEHYKFTSSEENTQKVDHYANSLQNLAKSLEELDKATTNEQATQVK--NQLENA	285	
QY	217	QKTIAD--TLIKODGRYVTFPKA-----SEIKNMPATTKDAQTNTIEPLDYWDS-	264	
Db	286	QKLKEMOPLIKET---NVKLYKAMESLEQVEKELKHNSLEANLIDLAKSKEIVREYEGK	342	
QY	265	-----WPODVRTGOVANWNGYQ---L	283	
Db	343	LNOSKNLPQLQLEBEAHSKLVQVEDFRKKFTSEQVTPKRVKRDLLAANNQOKIEL	402	
QY	284	VIAMMIPNQNDNHLYLLYKNYGONELSHKNVGPITPGYNSTAVSGWSGSAVLNSDSI	343	
b	403	TVSPENITVEEGEDVKFTVTAKOSDKTT---LDFSLLTKYNPSVRSIRSTNYKTNTDN--	458	
QY	344	QLFYRTVDTSDNNTNHOKIASATLYLTDNNGNYSLAQVRNDIYVEBGDYYYQTQDWKA	403	
b	459	---HKIAETIKNLKLNESQVTLLKAKDSSGNV---VEKFTTIV-----QKKE	501	
QY	404	TNKGADNIAMRDAAHIE-----DNGDRYLVFEASTGLYNGEDQIYNWLNLYCGD	454	

Qy 6 KNWVVVTLSTAALVFGATTVNASADTNIENNDSST-----VQVTTGDNDIAVKSVTLSG 59

```

Db 5 KDDSVKTEVAAPY-----PSMAQTDOGNSSSELETTKMEIPTDIDKAVEPEKTA 59
Qy 60 GOVSAASDITRTSA-----NANSASSAANTONSQVASSAAITSTSSAASL----- 108
Db 60 GETS-ATDTGKREKQLOOWKNNLNKNDVNTILSHEQKNEFKTIDETNDSALLELENOF 118
Qy 109 NNT-----DSKAAQENTNTAKNDTQKAAAPANESSEAKNEPAVNVNDSSAA 154
Db 119 NETNRLHLIKQHEVEKDKAKQO---KTLKQSDT-KVDLSNIDKE-----LNHOKSQVE 169
Qy 155 KNDQOQSSKNTAKLNKDAENVVKKAGIDPNLSLTDQIKALNKMNFSAKASGTQMTYN 214
Db 170 KMAEQGITNEDKSLMKIEDIRKQA-----QOADKEDAEVKVQLEBEAHSKLVQVE 224
Qy 215 DFQKIADTLIKQDGRYTVPPFKASE-----IKNPAATTQDAQ---TNTIEPLDQVND 263
Db 225 DFRK-----KFTSEQVTPKRVKRDLAANENNOOKIELTVSPENITVYE 269
Qy 264 SWPVQDVRTGOVANWNGYQLVIAMGIPNDNHLYL-----LYNKYGDNELSHWKNVGP 319
Db 270 G---EDVK-----FTVTAKSDSKTTLDFSDLLTKY----- 296
Qy 320 FGYNSTAVSQWSSGSAVLNSDINSIOLEFTRVDTSDNNTNHOKIASATLYLTDNNGNVS 379
Db 297 ---NPSVSDRISTNYKTNDN-----HKIAETIKNLKLNESQVTLKAKDDSGNV--- 344
Qy 380 QVRNDYIVFEGDGYIYQTYDQWKATNKGADNIAMDAHVIE-----DNGDRYLVF 430
Db 345 -VEKTFITV-----OKKEEKQVPTPEQKDSKTEEKVQPEKSNKNOQLQELIK 393
Qy 431 EASTGLENYQGEDQIYNWLNNGYGGDDAFNKLFRILSNDDIKS-----RATWAN--- 479
Db 394 SAQOELEKLE-----KAIKELMEQPEIPSPNPEYGIQKSIWESQKEP 434
Qy 480 ---RAIGILKLNKDEKNPKVAELYSPLISAPWSDTERP-----NVVGLGNKYILFAA 530
Db 531 TRLRGSNDAMNANTAVGDNVAVGVADSLTGSYKPL-----NDSGVVLTSAPVANWR 586
Db 495 IKKTFESDMKTKEDNYGSLNDALKYFKEYFTLPENKIKQIYDLDLKKVQDQPA--- 551
Qy 587 TATYSYAVPVAGKDDQ-----VLVTXYTNRNGVAGKWDSTWAPSLQINPDNTTT 640
Db 552 -----PIPENSEMDQAKKATVASYMS-----KVLGDGVHQ-----HLQKKNSK 592
Qy 641 VLAKWTQGDWINDSSENLDMIGDLSAALPGERDKPVDWDLIG-----YGLK- 689
Db 593 IV-----DLFKELEAIKQOTIFIDNAKTEVIDNLVH-DAFSKMNATVAKFQKLET 644
Qy 690 --PHDPATPNDPETPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 747
Db 645 NTPETPTPKIPELQAPDTPQAPDTPHPVESPESKAPVPSPTPTPEAPHVPSKAP 704
Qy 748 ETOTNNRLPOT 758
Db 705 EAPRVPSPKT 715

```

RESULT 10

```

US-08-714-481-2
; Sequence 2, Application US/08714481
; Patent No. 5766606
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5766606-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville

```

```

; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,481
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-6100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-714-481-2

```

```

Query Match 5.98; Score 242; DB 1; Length 984;
Best Local Similarity 20.78; Pred. No. 3.5e-09;
Matches 176; Conservative 112; Mismatches 325; Indels 238; Gaps 34;

```

```

Qy 6 KNAVVTLSAALVFGATTYNASADTNIENNDST-----VOVTTGDNDAIVKSVTLGS 59
Db 5 KDDSVKTEVAAPY-----PSMAQTDOGNSSSELETTKMEIPTDIDKAVEPEKTA 59
Qy 60 GOVSAASDITRTSA-----NANSASSAANTONSQVASSAAITSTSSAASL----- 108
Db 60 GETS-ATDTGKREKQLOOWKNNLNKNDVNTILSHEQKNEFKTIDETNDSALLELENOF 118
Qy 109 NNT-----DSKAAQENTNTAKNDTQKAAAPANESSEAKNEPAVNVNDSSAA 154
Db 119 NETNRLHLIKQHEVEKDKAKQO---KTLKQSDT-KVDLSNIDKE-----LNHOKSQVE 169
Qy 155 KNDQOQSSKNTAKLNKDAENVVKKAGIDPNLSLTDQIKALNKMNFSAKASGTQMTYN 214
Db 170 KMAEQGITNEDKSLMKIEDIRKQA-----QOADKEDAEVKVQLEBEAHSKLVQVE 224
Qy 215 DFQKIADTLIKQDGRYTVPPFKASE-----IKNPAATTQDAQ---TNTIEPLDQVND 263
Db 225 DFRK-----KFTSEQVTPKRVKRDLAANENNOOKIELTVSPENITVYE 269
Qy 264 SWPVQDVRTGOVANWNGYQLVIAMGIPNDNHLYL-----LYNKYGDNELSHWKNVGP 319
Db 270 G---EDVK-----FTVTAKSDSKTTLDFSDLLTKY----- 296
Qy 320 FGYNSTAVSQWSSGSAVLNSDINSIOLEFTRVDTSDNNTNHOKIASATLYLTDNNGNVS 379
Db 297 ---NPSVSDRISTNYKTNDN-----HKIAETIKNLKLNESQVTLKAKDDSGNV--- 344
Qy 380 QVRNDYIVFEGDGYIYQTYDQWKATNKGADNIAMDAHVIE-----DNGDRYLVF 430
Db 345 -VEKTFITV-----OKKEEKQVPTPEQKDSKTEEKVQPEKSNKNOQLQELIK 393
Qy 431 EASTGLENYQGEDQIYNWLNNGYGGDDAFNKLFRILSNDDIKS-----RATWAN--- 479
Db 394 SAQOELEKLE-----KAIKELMEQPEIPSPNPEYGIQKSIWESQKEP 434
Qy 480 ---RAIGILKLNKDEKNPKVAELYSPLISAPWSDTERP-----NVVGLGNKYILFAA 530

```


;; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
;; NUMBER OF SEQUENCES: 34
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
;; STREET: 1100 New York Avenue, N.W., Suite 600
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; FILING DATE: 05-SEP-1997
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/024,707
;; FILING DATE: 06-SEP-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Esmond, Robert W.
;; REGISTRATION NUMBER: 32,893
;; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; TYPE: amino acids
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-923-992A-2

Query Match 5.88; Score 241; DB 4; Length 1164;
Best Local Similarity 20.08; Pred. No. 5.3e-09;
Matches 191; Conservative 130; Mismatches 347; Indels 288; Gaps 40;

QY 10 VVTLSTALVFGA-----TTVNAS-----ADTNINNDST-----VOV 43
DB 21 VASVAVASLPMGSAHASELVKDDSVKTEVAAPKYPNSAQTDOGNSSSELETTKMEI 80
QY 44 TTGNDIAVKSVTILGSGVSAASDTHRTSA-----NANSASSAANTQNSQVASSAAIT 99
DB 81 PTTDIKAVPEVETAGETS-ATDTGKREKQLQKWKNNLNKNDVDNTILSHEQKNEFKTKI 139
QY 100 SSTSSRAASL-----NNT-----DSKAAQENTNTAKNDTQ-----129
DB 140 DETNDSALLELENOFNETNRLIHIKQHEVEKDKKAKQO---KTLKQSDTKVDLSNIDKE 197
QY 130 ---KAAPANESSEAK---NEPAVNVNDSSAANKND---QSSKK-----164
DB 198 LNHQKSOVEKMAEQGKITNED---KDSMLKIEDIRKQAOQADKEDAEVKEVREELGKL 253
QY 165 --NTTAKLAKDAENVVK-----AGID---PNSLDDQIKALKNNKFSKAASKGTOM 211
DB 254 FSTTRAGLQDQETQEHVKKETSSEENTQKVDEHYANSLQNLAKLSLELDKATTTNEQATOV 313
QY 212 TYNDF-----QKTD--TLIKODGRYTPFFKA-----SEIKNPAATTKDAQNTTI 256
DB 314 K-NQLENAQKLKEIQPLIKET---NVKLYRAMSESLEQVEKELKHNSLEANLVLAKSK 369
QY 257 EPLDWDS-----WPVQDVRTGOVAN 277
DB 370 ELVRETEGKLNQSKNLPKOLEEEAHSKLQVVEDEFKFKTSEQVTPKRVKRDLAAN 429
QY 278 WNGYQ---LVIAMGIPNDNDHLYLKYKGNELSHWKNVGFYGNSTAVSOWSGS 334
DB 430 ENNOQKIELTSPENITVYGEDVKFTVTAKSDSKTT--LDFFSLLTKYNPVSVDRISTN 487

QY 335 AVLNSDNTSLQFTRVDTSDNNTNHQKIASATLYLTLDNNGVSLAQNRYNDYVIFEGDYY 394
DB 488 YKNTNDN-----HKTAETITKLNKLNESQTVTLAKADDSGNV-----VEKFTTIV 533
QY 395 YQTYDQWKATNGAGDNIAMRDHVI-----DNGDRYLVFEASTGLNYSQGEDQI 445
DB 534 -----QKKEEKQVPKTPQKDSKTEEKVQEPKSNKNDKNOLELIKSAQOELEKLE-----583
QY 446 YNNLNYGGDAPNFKSLFRILSNDDIKS-----RATWAN-----AAIGILKLNKDE 491
DB 584 -----KAELMEQEPISNPEYGIQKISWESQKEPQEQEATISFKKIIGDS 629
QY 492 KNPKVAELYSPLISAPMVSDIERP-----NVKLGKYYLFAATRLNNGSNDAMWA 545
DB 630 SSKYVTEHYFNKYKSDFNQOLHAQOMELFRKVQVQVNMKYPDNAEIKKIFESDMKRTKD 689
QY 546 NYAVGDNVAMGVYVADSLTGSYKPL-----NDSGVVLTASVPANMRTATYSYAVPVAGKD 601
DB 690 NYGSLNDALKGFEKYFLTPFNKIKQIVDDLDKKVQEQDPA-----PIPENSEM 739
QY 602 DQ-----VLVTSYMTNRRNGVAGKGMDSWAPSFLLQINPDNTTTLAKMTNQGDIWDD 655
DB 740 DQAKEKAKIAVSKYMS-----KVLGGVHQ-----HLQKKNSKIV-----DLFKEL 780
QY 656 SSENLDMIGLDSAAALPGERKRPVWDLLIG-----YGLK---PHDPATNDPPTP 702
DB 781 EAIKQGTIFIDNAKTEVEIDNLVH--DAFSKMNATVAKFQKLENTTETPTDKPIELP 839
QY 703 TTPETPETNTPKTKTPENPGTPPTPNTPTNPTPELTPETPKQPTQTNNRLPOT 758
DB 840 QAPDTPQADPTPHVPESPKAPEAPRVPSKPTPEAPHVPSKPAPEAPRVPSPKT 895

RESULT 13
US-08-923-992A-8
; Sequence 8, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-8

```

Query Match	5.8%;	Score	240.5;	DB	4;	Length	1098;
Best Local Similarity	19.2%;	Pred.	No.	5.3e-09;			
Matches	179;	Conservative	132;	Mismatches	360;	Indels	259;
Gaps	34;						
Qy	6	KNWAVVLTSTAALVGAITVNASADTNTENDDST	-----	QVTTGDNDAIAKSVTLGS	59		
Db	11	KDQSVKTYEAAKPY	-----	PSMAQTDOGNSSSELETTTWEITPTDIKKAPEVPEKTA	65		
Qy	60	GOVSAAS	-----	DTTIRTSANANSASSAANTQNSNOVASSAITS	100		
Db	66	GETSATHTGKREKQLQWKNLKNVDVNTILSHQKNEFKTIDETNDSDALLENQFN	125				
Qy	101	STFSSAASLN	-----	NTOSKAAQENT	-----	NTAKNDOTQKAAPNESSEAKNEP	144
Db	126	ETNRLLLHKEHVEVEKKAKQOQKTLKOSDTKVDSLIDIKELNHQKSEQAGITNEDKDSM	185				
Qy	145	AVNVNDSSAAKNDQOQSKK	-----	NTAKLNKDAENVYKK	-----	180	
Db	186	LKKIED	-----	IRKQAQPDKKEDAEVYREELGLKLFSTKAGLDOEIQEHVYKKTSSBENT	242		
Qy	181	AGID	-----	PNSLTDOOIKALKNMFSSKAAGSGTQMYNDP	-----	OKIAD	-----
Db	243	QKVDHANSJONLAQSKSLBELQKATTNEQATQVK	NOFLENAQKLEIQIPLIKET	-----	N	298	
Qy	232	VPFPKA	-----	SIKKNPAAATTKDAQNTNIEPLDVWDS	-----	264	
Db	299	VKLYKAMSESELEQVEKELKHNSLEANLDAKSKVIREYEGKLNQSKNLPULKQLEEA	358				
Qy	265	-----	HPVDQVRCQVANWNGYQ	-----	LVIAMGIPNQNNDHIYL	300	
Db	359	HSKLQVQVEDFRKFKTSEQVTPKKRLKRDLAANENNOQKIETVSPENITVVEGEDVKF	418				
Qy	301	LYNKYGDELHWNKNGPIFGYNSTAVSQWGSAGVLSNDNSITQLPYTRVDTSNNTNHQ	360				
Db	419	TVTAKSDKTT	-----	LDPSDLITTKYNPSVDRISNTYKNTDN	-----	HKTAETITKLNKLN	471
Qy	361	KIASATLYLTDNNGVNSLAQVRNDYIVFEGDGYTYQYQWKATKNAGDNIAMRDAHVIE	420				
Db	472	ESQTVTLKAKDGSNV	-----	VEKTFITV	-----	OKKEEQVQPKTPBQKDSKTEE	517
Qy	421	-----	DGNGDRYLVEASTGLENGYGEDQIYNWLNLYGGDDAFNTKSLFRILSNDDI	471			
Db	518	KVPOEPKSDKNQLOBLIKSAQOQLEKL	-----	KAKELMEQPEI	558		
Qy	472	KS	-----	RATWAN	-----	AATGILKLNKDEKNPKVAELYSPLISAPMVSDIERP	516
Db	559	PSNPEYGIQKSWESQKEPIQEAITSFRRKIIGDSSSKYTYEHTFYFNKYKSHFMYQLHAQM	618				
Qy	517	-----	NVWKLGNKYFLAATRLNRGNSDDAMNANYAGVDNVMVGYVADSLTSGYKPL	570			
Db	619	EMLTRKVVQYMKNYPDNAEIKKIFESDMKRTKEDNGYSLNDALGKYPEKFLFPFNKIK	678				
Qy	571	-----	NDSGVVLTASVPANRRTATYSIYAVPVAGKDDQ	-----	VLVTSYMTNRNGVAGKGM	621	
Db	679	QIVDDFDKKVEQDQA	-----	PIPENEMDQAKEAKIAVSKYMS	-----	KVL	721
Qy	622	DSTWAPSFELQINPDNTTVTLAKMTNOCGDWTWDDSSENLMDIGDLSAALPGERDKPVDW	681				
Db	722	DGVHO	-----	HLQKNHISKIV	-----	DLFKELEAIKQOTIFDIDNAKTEVELDNLVH	768
Qy	682	DLIG	-----	YGLK	-----	PHDPATPNQDETPTTPTETPNTPTKPTKPTPENPGPTQT	728
Db	769	DAFSKKNATVAKFQGLTNTPTPDTPKIPELPQAPDTPQADTPHVPQSPKAPEAPRV	828				
Qy	729	PNTPTPTPEIPLTPTPTKQPEPTQTNRLPQT	758				
Db	829	PESPCTPEAPHVPSKPAPEAPRVPEPSPKT	858				

RESULT 14
 US-08-923-992A-4
 ; Sequence 4, Application US/08923992A
 ; Patent No. 6280738
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
 ; TITLE OF INVENTION: Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,992A
 ; FILING DATE: 05-SEP-1997
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/024,707
 ; FILING DATE: 06-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Esmond, Robert W.
 ; REGISTRATION NUMBER: 32,893
 ; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1104 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-923-992A-4

Query Match	5.8%	Score 238;	DB 4;	Length 1104;
Best Local Similarity	19.6%;	Pred. NO. 8.1e-09;		
Matches 185;	Conservative 129;	Mismatches 351;	Indels 278;	Gaps 37;

Qy	6	KNAVVTLSTAALVFGATTNASADTTNIENDST-----VQVTTGDNDIAVKSVTLGS	59
Dd	11	KDSVKTEVAAPY-----PSMAQTDOGNSSSELETTTRMBIPTHIKKAYPEVEKTA	65
Qy	60	GQVSAAAS-----DTTIRTSANANSASSAANTQNSNQVASSAAITS	100
Dd	66	GETSATDTGKREKLOQWKNLKNKYDNTLTLSHEQNFEFKTKIDETNDSDALLELNQFN	125
Qy	101	ST-----SSAASLNNTDTSKAAQENTNTAKNDDTQKAAPANESSEARN	142
Dd	126	ETNRLLHIKOHEEVEKHNPNOOKTLKQSDTVLDLSNIDELNHQKSQVEAMAEQAIGTN	185
Qy	143	EPAVNVDSSAAKND-----QSSSKNTTAKLUNKDAENVVK-----KAGID----	184
Dd	186	ED-----KDSMLKIETDIRKOAQQADKK-----EDAEVKVREELKLFSSSTKAGLDQOI	234
Qy	185	-----PNSLTDDOIKAALKNFESKAAKSCOTWTYNDF-----QKI	219
Dd	235	QEHVKETSEEONTQKVDEHYANSLQNLAQSKLEELDKATTTNEQATOVK-NQFLENAOKL	293
Qy	220	AD--TLIKODGRYTPPFKA-----SETKNPPAATTKDAQTNITIEPLDVWDS-----	264

[illegible]

RESULT 15
 US-07-638-431-2
 : Sequence 2, Application US/07638431
 : Patent No. 5198535
 :
 : GENERAL INFORMATION:
 : APPLICANT: Hoffman, Stephen L.
 : APPLICANT: Charoenvit, Yupin
 : APPLICANT: Hedstrom, Richard
 : APPLICANT: Khumsmith, Srisin
 : APPLICANT: Rogers IV, William O.
 :
 : TITLE OF INVENTION: Protective malaria sporozoite surface protein
 :
 : NUMBER OF SEQUENCES: 2
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: A. David Sperack
 : STREET: NMRDC Building 1 T-12 National Naval
 : STREET: Medical Center
 : CITY: Bethesda
 : STATE: MD
 : COUNTRY: USA
 : ZIP: 20814-5044
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.24
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/07/638,431
 : FILING DATE: 19910110
 : CLASSIFICATION: 424

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Spevack, Avrom D.
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 295-6759
: TELEFAX: (301) 295-4033
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 826 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-638-431-2

```

Query Match 5.3%; Score 220; DB 1; Length 826;
Best Local Similarity 20.7%; Pred. No. 1.1e-07;
Matches 87; Conservative 46; Mismatches 132; Indels 156; Gaps 17.

415	QY	DAHVTEGNGDRYLVEASTGLENTQGE-----DOIYNWLNKGGDDAFNIKSLPRLISND	469
		: :	
43	Db	DIHILLDGG-----SIGYSNNKAHVIPMLNTLVDNLNINSDEINVSLTFSNRSRE	94
		: :	
470	QY	DIKSRAIWANAAGILKLNKDEKNPKVAEL-----YSP-----LISAPWVSDETERP	516
		: :	
95	Db	LIKLE-----GYGTSKSDSLFLIAHLQNNYSPNGNTNLTSALLVVDVTLINERYRP	146
		: :	
517	QY	NVYKLGNNKYFLFAATRLNRGSDNDAMNANYAGD-----NVAMGYVADSLTGSYK-	568
		: :	
147	Db	DAIOL-----AIIITDGIPIND-----LPRSTAVVHQLKRRKHVNAIIG-VGAGVNNENYR	195
		: :	
569	QY	-----PLNDSGVLTASVAPANWRATYTYVYAVPVAGKDDOVLVTSYMTNRNGVA	617
		: :	
196	Db	ILVGCDRYAPCPYYSSG-----SWNEA-----QNWIKFELT-----	226
		: :	
618	QY	KGKMDSTWAPSFLLQINPDNTTVLAKMTNQGDW-IWDDSSENLD-----	661
		: :	
227	Db	-----KVCOEVERIAHCWKWSEWSECSTTCDEGRKIRRRQILHPGC	267
		: :	
662	Yy	-----MTGDLSAAL-----PGSRDRKPVWDMLIGYGLKPHOPATPNDETP	702
		: :	
268	Db	VSEMTTPCKVRDCPQIPIPPVVPNKIPEKSPNPEPVN-----PNPDNPNNPNPNP	318
		: :	
703	Yy	TTPTETPTNPTKPTKPTPNBGTQPTNPTNPTPEIPLTPKQETQTNRLRLPQTGNA	762
		: :	
319	Db	NNP	378
		: :	
763	Yy	N 763	
379	N	379 N 379	

Search completed: September 21, 2002, 23:37:39
Job time: 5080 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 23:32:00 ; Search time 7643.51 Seconds
(without alignments)
4179.663 Million cell updates/sec

Title: US-09-995-587A-2
Perfect score: 2367
Sequence: 1 atgtataaaacggtataaaa.....ttaacaaacgctgatttaac 2367

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.2	3.4	989	12 CNS02HA4	AL197365 Tetraodon
2	76.2	3.2	501	12 CNS048173	AL444958 Fugu rubr
3	75.4	3.2	735	12 CNS04NSM	AL299119 Tetraodon
4	74.8	3.2	773	12 CNS01VTG	AL169549 Tetraodon
5	72.6	3.1	494	12 CNS048073	AL444858 Fugu rubr
6	71.8	3.0	450	12 CNS025683	AL018519 Fugu rubr
7	68.2	2.9	470	12 CNS018463	AL011359 Fugu rubr
8	67.2	2.8	806	12 CNS006944	290754 F. rubripes
9	67.2	2.8	816	12 CNS04AEE	AL281759 Tetraodon
10	65.6	2.8	959	12 CNS006555	AL062806 Drosophila
11	64	2.7	619	12 CNS0047503	AL444288 Fugu rubr
12	63.6	2.7	698	10 BM358689	BM358689 GA_Ea001
13	63.6	2.7	979	12 CNS06PDW	AL409242 T7 end of
14	63.4	2.7	619	12 CNS06PDW	AL444386 Fugu rubr
15	63.2	2.7	641	12 CNS047601	AQ946120 Sheared D
16	62.8	2.7	1007	12 CNS06X9S	AL419462 T3 end of
17	62.6	2.6	515	10 C90898	C90898 C90898 Dict

18	62.6	2.6	925	10 BG441163	10 BG441163 GA_Ea001
19	62	2.6	619	12 FR0013713	AL004959 F. rubripes
20	61.8	2.6	550	12 FR0043207	AL130699 Fugu rubr
21	61.6	2.6	824	12 AZ183454	AZ183454 SP_1005_A
22	60.8	2.6	519	10 BG226324	BG226324 KQ20C02_Y
23	60.8	2.6	919	12 BH138523	BH138523 ENT0807TR
24	60	2.5	479	12 BH211305	BH211305 Sml-50G22
25	60	2.5	933	12 AZ204694	AZ204694 SP_0100_A
26	59.8	2.5	534	12 BH203231	BH203231 Sml-57E13
27	59.6	2.5	394	9 AU179547	AU179547 AU179547
28	59.4	2.5	257	12 FR0026883	AL019712 F. rubripes
29	59.4	2.5	273	12 FR0047466	AL444251 Fugu rubr
30	58.4	2.5	836	12 AQ411961	AQ411961 CPG0931A
31	59.4	2.5	1101	12 CNS0000CT	AL065406 Drosophila
32	59.2	2.5	607	9 BE124614	BE124614 EST393649
33	59.2	2.5	687	10 BG447645	BG447645 NF003A12S
34	58.4	2.5	893	12 AZ542140	AZ542140 ENTEN36TF
35	58.4	2.5	865	12 AZ685375	AZ685375 ENTKV53TF
36	58.2	2.5	453	9 AU179269	AU179269 AU179269
37	58.2	2.5	610	9 AU180091	AU180091 AU180091
38	58.2	2.5	567	12 FR0006951	AU180168 AU180168
39	57.8	2.4	567	12 CNS000880	290761 F. rubripes
40	57.6	2.4	593	12 CNS000880	AL051540 Drosophila
41	57.2	2.4	326	9 AI443192	AI443192 sa45f05.Y
42	57.2	2.4	454	9 AI496419	AI496419 sb04e01.Y
43	57.2	2.4	474	10 BF595475	BF595475 su77f03.Y
44	57.2	2.4	488	9 BE020655	BE020655 sm51n06.Y
45	56.8	2.4	748	12 AZ183618	AZ183618 SP_1001_A

ALIGNMENTS

RESULT 1
CNS02HA4
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence P7 end of clone
139021 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL197365
VERSION
AL197365.1 GI:7835515
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE
1. (bases 1 to 989)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL
Unpublished
REFERENCE
2. (bases 1 to 989)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL
Unpublished
REFERENCE
3. (bases 1 to 989)
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
location/Qualifiers
1. .989
/organism="Tetraodon nigroviridis"

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
1 (bases 1 to 470)
Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrانيا, Y.,
Williams, G. and Brenner, S.
Direct Submission
Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk
V type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES
source

Location/Qualifiers
1..470
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 016E10"
/clone="016E10Ac6"
BASE COUNT 149 a 79 c 150 g 65 t 27 others
ORIGIN

Query Match 2.9%; Score 68.2; DB 12; Length 470;
Best Local Similarity 48.6%; Pred. No. 0.0001;
Matches 169; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 127 gttacacaggtgataatgattgctgttaaaagtgtgacactgttagtggtcaagtt 186
DB 408 GCTACTACTGCTGCTACTACTGCTGCTACTACTGCTGCTACTACTGCTGCT 349
QY 167 agtcagctagatagacactattgaaactcttctgctgaatgcaaatgctctctgccc 246
DB 348 ACTAATGCTGCTGCTACTACTGCTGCTACTACTGCTGCTACTACTGCTGCT 289
QY 247 gctaatcacaaaattctaacagcaagtagcaagttctgctgcaataacatcatcata 306
DB 288 GCTACTACTGCTGCTACTGCTGCTGCTACTACTGCTGCTACTACTGCTGCT 229
QY 307 agttccgagcttcttaataacacagatagtaaaaggcgtcaaaataatacata 366
DB 228 NCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 169
QY 367 gccaaaaatgatgacacgcaaaaggcgtcaaaaggcgtcaaaaggcgtcaaaaggcgt 426
DB 168 ACTGCTACTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 109
QY 427 gaaccgctgtaaacgttaattgattcttcagctgcaaaaatgatgat 474
DB 108 ACTACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 61

RESULT 8
FR0006944
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

F.rubripes GSS sequence, clone 133B16Ac7, genomic survey sequence.
290754
290754.1 GI:1867968
GSS; genome survey sequence.
Takifugu rubripes.
Takifugu rubripes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 619)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrانيا, Y.,
Williams, G. and Brenner, S.
Direct Submission
Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk
Vector: pBluescript II KS

V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES
source

Location/Qualifiers
1..619
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 133B16"
/clone="133B16Ac7"
BASE COUNT 75 a 150 c 116 g 175 t 103 others
ORIGIN

Query Match 2.8%; Score 67.2; DB 12; Length 619;
Best Local Similarity 48.3%; Pred. No. 0.00018;
Matches 174; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 63 tgcacaaactgtaaatgcatccgcggcacaaatattgaaacaaatgattcttctactgt 122
DB 14 TGCTACTANNACTGCTGCTACTACNGCTNCTGCTACTTAATGCTACTGCTGCTAC 73
QY 123 acaagttacacaggtgataatgattgctgtttaaaagtgtgacactgttagtggtca 182
DB 74 TGCNACTACTGCTACTGCTGCTACTACTGCTACTGCTGCTGCTGCTGCTGCTGCT 133
QY 183 agttagtcagctagtagacactattgaaactcttctgctgaatgctgcaataacatcata 302
DB 134 TGCTGCTGCTGCTGCTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 193
QY 243 tgcgcgttaatacacaaaattctaacagtagcaagttctgctgcaataacatcata 362
DB 194 TACTACTACTGCTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 253
QY 303 tacaagttccgagcttcttaataacacagatagtaaaaggcgtcaaaaggcgtcaaaaggcgt 422
DB 254 TACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 313
QY 363 tacagccaaaatgatgacacgcaaaaggcgtcaaaaggcgtcaaaaggcgtcaaaaggcgt 422
DB 314 TACTGCTACTGCTGCTGCTACTANTGNTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 373

RESULT 9
CNS04ABE
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Tetraodon nigroviridis genome survey sequence T7 end of clone
095G24 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL281759
AL281759.1 GI:8020086
GSS; genome survey sequence.
Tetraodon nigroviridis.
Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 806)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizes, C., Fisher, C.,
Bounau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 806)
Roest-Crolius, H., Jaillon, O., Dasilva, C., Bounau, L., Fisher, C.,
Bernot, A., Fizes, C., Wincker, P., Brothier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished

```

REFERENCE 3 (bases 1 to 806)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   This sequence is a single read and was generated as part of a large
COMMENT   scale clone-end sequencing project of the Tetraodon nigroviridis
          genome. For more information, please take a look at
          http://www.genoscope.cns.fr/tetraodon.
FEATURES  Location/Qualifiers
            1..806
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone="095G24"
            /clone.lib="G"
            /note="Genoscope sequence ID : COR095BD12LPI-end : T7"
BASE COUNT 247 a 87 c 99 g 333 t 40 others
ORIGIN
Oy 874 accaacaatgatacatctcttcttataatgaatggtgataatgaatgaattgaatt 933
Db 737 AATAAATAGTAATAATCATAAATAAATAAATAAATAAATAAATAAATAAATAA 678
Oy 934 cattggaagaatgaggtcccaatttttggtctataattctaccgcggtttccacaagaattg 993
Db 677 AATAGTAATAATGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 618
Oy 994 tcaggatcagctgtttgaacagtgataactctcaatttatttatacaagggtagac 1053
Db 617 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 558
Oy 1054 acgtctgataacaataccaatcatcaaaaaattgctagcgtctactctttttaaactgat 1113
Db 557 AAMSTATTATATACAAAHMACMAAMCMCMCMCMCMCMCMCMCMCMCMCMCMCM 498
Oy 1114 aataatggaatgtatcactcgtcaggtacgaaatgactatattgttgaaggtgat 1173
Db 497 AATAAATAAATGTAATAATAATAATAATAATAATAATAATAATAATAATAATA 438
Oy 1174 ggcattactaccacattatgataatggaagcgtactaacaagggtccgataatt 1233
Db 437 AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 378
Oy 1234 gcaatgcgtgatgctcatgtaattgaagatggttaattggtgacgtacatttgaat 1293
Db 377 GGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 318
Oy 1294 gcaagtactggttggaaattatcaaggcggaggaccaaatttataactggttaattat 1353
Db 317 AATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 258
Oy 1354 ggcggagatgacgattatcaatcaagcgttattagaattctttcccaatgatatt 1413
Db 257 AGTAATGATAATCATAGTAATAATAATAATAATAATAATAATAATAATAATA 198
Oy 1414 aa 1415
Db 197 AA 196

RESULT 10
LOCUS   CNS00655
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
          BAC13F18 of RPCI-98 library from Drosophila melanogaster (fruit
          fly), genomic survey sequence.
ACCESSION AL062806
VERSION   AL062806.1 GI:4943996
KEYWORDS  GSS.

```

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 959)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers
1..959
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone.lib="RPCI-98"
/clone="BAC13F18"
/note="end : TET3"

BASE COUNT 525 a 63 c 52 g 106 t 213 others
ORIGIN

Query Match 2.8%; Score 65.6; DB 12; Length 959;
Best Local Similarity 31.9%; Pred. No. 0.00046;
Matches 210; Conservative 111; Mismatches 337; Indels 0; Gaps 0;

Oy 291 aataacatcatctacaagttccgcagcttcattataaacaacagatagtaacggctca 350
Db 298 AWAAMWMAAAMAAAAAAMAAACACAAAMMAAAMAAAAAAMAAAAAACAACACAC 357
Oy 351 agaaataactaatacagcacaagaatgatgacgcaaaagctgcacagctcaagcattc 410
Db 358 AMAWTTAAAAACACGCGSSSYMMAMMMMMMMCMMAAASRTTSTTSCCMTCCCCCC 417
Oy 411 ttctgaagctaaatgaaccagctgtaaacggttaattgattcttcagctgcaaaaatga 470
Db 418 CCCCSARAAARAAAAAATTTTGTGMMKTTKWTATGTAATGTAATGTAATGTAATGTA 477
Oy 471 tgatcaacaatccctgtaaaagaatactacgcgttaagttaacaaggatgctgaaacgt 530
Db 478 RARGAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAA 537
Oy 531 tgtaaaaagcggggaattgatctcaacagtttaactgatgaccagatgaagcattaaa 590
Db 538 ARWAKDAARAGGGGAARAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAA 597
Oy 591 taagatgaactctcgaaagctgcaagctcgtggtacacaaatgacttataatgattcca 650
Db 598 AAAAAAARAAAAAGAAWAAWRAAKDAAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAA 657
Oy 651 aaagatgctgatcgttaatacaacaagatggtcggtacacagctccattctttaaagc 710
Db 658 AAAMWTTTATTTAAWTTAAKRAAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAA 717
Oy 711 aagtgaaatcaaaaatgctgcgcgtacactaaagatgcacaaactaactattga 770
Db 718 RDAARRAAAAAAMWADGRKAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAADGD 777

Saccharomyces exiguus, genomic survey sequence.

AL409242.1 GI:12176588

Accession AL409242.1 GI:12176588

Version AL409242.1 GI:12176588

Keywords Saccharomyces exiguus

Source Saccharomyces exiguus

Organism Saccharomyces exiguus

Reference Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

Authors Bon, E., Neuveglise, C., Lepingle, A., Wincker, P., Artiguenave, F., Gaillardin, C. and Casaregola, S.

Title Genomic exploration of the hemiascomycetous yeasts: 6.

Journal FEBS Lett. 487 (1), 42-46 (2000)

Medline 20584716

Reference 2 (bases 1 to 979)

Authors Souclet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Liorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.

Title Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

Journal FEBS Lett. 487 (1), 3-12 (2000)

Medline 20584711

Reference 3 (bases 1 to 979)

Authors Genoscope.

Title Direct Submission

Journal Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

Comment This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

Features source Location/Qualifiers

1. 979

/organism="Saccharomyces exiguus"

/strain="CBS 379"

/db_xref="taxon:34358"

/clone_lib="AV0AA013G05"

/clone="AV0AA013G05"

/note="end : T7"

BASE COUNT 226 a 205 c 217 g 331 t

ORIGIN

Query Match 2.7%; Score 63.6; DB 12; Length 979;

Best Local Similarity 49.1%; Pred. NO. 0.0013;

Matches 168; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 55 gtatttggtgcaacaactgtaaatgcatccgagacacaaatattgaaacaatgattct 114

DB 408 GGAACCTGGTCACTGATGGTAAATGGTTCTCAAACTACCAACCGATGGTAATGGTCTCTA 349

QY 115 tctactgtacaagtacacaggtgataatgattgctgtttaaagtgtagcacttgg 174

DB 348 ACAACACTACAGGAACCTGGTGAACCGATGGTAAATGGTTCTCAAGTACAACTGATGGT 289

QY 175 agtgggtcaagttagtcacgtgatacagcattatagaactcttgcgttaagtcaaatag 234

DB 288 AATGGTCTTTAACTCAACTGATGGTAACTGATGGTCAAACTACAACTGATGGTAAATGGT 229

QY 235 gcttctcttcgctgatacacaacaattcttaacagtcgaagtcgaagttctgctgcaata 294

DB 228 TCTCTAACTGCAACCGATGGTAAACAGTGTCAAACTACCACTACAGGAACCTGCTACAAT 169

QY 295 acatcatctacaagttccgcagcttcattataaataacacagatagtaaaagcggtcaagaa 354

DB 168 ACCGATGCAACAGATGGTAACGGTTCCCAACAACAGATGATGGTAATGGTACACATACG 109

QY 355 aatactaatacagccaaataatgatgacacgcgcaaaaagctgca 396

DB 108 ACAACTAATGGTGGTGTATGATGGTAATGAGATGGATCATCATCA 67

RESULT 14

LOCUS FR0047601

DEFINITION Fugu rubripes GSS sequence, clone 217A131G11, genomic survey sequence.

ACCESSION AL444386

VERSION AL444386.1 GI:12052222

KEYWORDS GSS: genome survey sequence.

SOURCE Takifugu rubripes.

ORGANISM Takifugu rubripes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformia; Tetraodontiformes; Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 619)

AUTHORS Clark, M.S.

TITLE Direct Submission

JOURNAL Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. UK Email: biohelp@hgmrc.ac.uk

COMMENT Vector: pBluescript II KS

V-type: phagemid

PRIMER: KS

DEFSCR:

One pass dye-terminator sequencing of BAC (pBelobACII) cloned genomic sequence

The BACs can be obtained from http://www.incyte.com.

Features source Location/Qualifiers

1. 619

/organism="Takifugu rubripes"

/db_xref="taxon:31033"

/clone="217A131G11"

/clone_lib="BAC 217A13"

BASE COUNT 102 a 172 c 92 g 177 t 76 others

ORIGIN

Query Match 2.7%; Score 63.4; DB 12; Length 619;

Best Local Similarity 47.1%; Pred. NO. 0.0013;

Matches 157; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 129 tacaacaggtgataatgattgctgttaaagttgacacttgtagtggtcaagttag 188

DB 9 TGCAACGACTACTACTACTGTTACTACTACTGCTGCTTCTACTACTACTACTACTAC 68

QY 189 tgcagctagtgatacagcattatagaactcttgcgttaagtcaaatagctcttctgcgc 248

DB 69 TGCTACTACTACTACTGCTTCTACTACTACTACTACTACTGCTGCTTCTACTACTGC 128

QY 249 taatacacaataattcaacagtcgaagtcgctgctgcaataacatcatcatacag 308

DB 129 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 188

QY 309 ttcgcagcttcattataaataacacagatagtaaaagcggtctcaagaaataactatacagc 368

DB 189 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAN 248

QY 369 caaaaatgatgacacgcgcaaaaagctgcaccagctaaacgaatctcttgaaagctaaaatga 428

DB 249 TACTACTGCTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 308

QY 429 accagctgtaaacgcttaattgattcttcagctgc 461

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 23:37:45 ; Search time 180.95 Seconds
(without alignments)
3213.122 Million cell updates/sec

Title: US-09-995-587A-2
Perfect score: 2367
Sequence: 1 atgtataaaagcgtaaaaa.....ttaacaaacgtcgatttaac 2367

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	80	3.4	10317	3	US-09-058-746-1
C 2	80	3.4	10317	4	US-09-438-142-1
3	68.8	2.9	3730	1	US-08-242-932-8
4	68.8	2.9	3730	1	US-08-714-481-8
5	68.8	2.9	3730	5	PCT-US95-06111-8
6	67.4	2.8	3305	3	US-09-068-043-1
7	64.2	2.7	3294	4	US-08-923-992A-7
8	64.2	2.7	3312	4	US-08-923-992A-3
9	64.2	2.7	3384	4	US-08-923-992A-5
10	64.2	2.7	3492	4	US-08-923-992A-9
11	64.2	2.7	4200	1	US-08-242-932-1
12	64.2	2.7	4200	1	US-08-714-481-1
13	64.2	2.7	4200	5	PCT-US95-06111-1
14	63.8	2.7	4200	4	US-08-923-992A-1
15	63.2	2.7	4873	1	US-07-638-431-1
16	63.2	2.7	4873	5	PCT-US92-00018-1
17	52	2.2	1365	2	US-08-870-827-4
18	52	2.2	2408	2	US-08-870-827-5
19	52	2.2	7218	1	US-08-232-463-14
20	51.8	2.2	3489	2	US-08-728-323A-1
21	51.8	2.2	32207	2	US-08-770-379-20
22	51.8	2.2	32207	4	US-08-757-669A-20
23	51.8	2.2	32207	4	US-09-230-371A-20
24	50.4	2.1	3666	2	US-08-682-517-13
25	50.4	2.1	3666	2	US-08-682-517-14
26	50.4	2.1	4197	2	US-08-682-517-7
27	50.4	2.1	4197	2	US-08-682-517-8

C 28 50 2.1 5102 1 US-08-494-168-1 Sequence 1, Appli
29 48.8 2.1 5163 3 US-08-700-651-1 Sequence 1, Appli
30 48.8 2.1 5163 3 US-08-928-361B-4 Sequence 4, Appli
31 48.8 2.1 5318 3 US-08-700-651-2 Sequence 2, Appli
32 48.8 2.1 5318 3 US-08-928-361B-3 Sequence 3, Appli
33 48.4 2.0 1079 1 US-07-781-355-1 Sequence 1, Appli
34 48 2.0 921 2 US-08-795-475-2 Sequence 2, Appli
35 48 2.0 1308 2 US-08-795-475-4 Sequence 1, Appli
36 47.6 2.0 3763 1 US-07-792-865D-1 Sequence 159, App
37 47 2.0 1924 3 US-08-961-083-159 Sequence 13, Appl
38 46.4 2.0 19124 2 US-08-487-826B-13 Sequence 14, Appl
39 46.2 2.0 1725 1 US-08-257-073-14 Sequence 10, Appli
40 46 1.9 1956 4 US-08-559-896B-1 Sequence 1, Appli
41 46 1.9 5181 1 US-08-257-073-10 Sequence 10, Appli
42 45.6 1.9 5511 3 US-08-928-361B-2 Sequence 2, Appli
43 45.6 1.9 7334 3 US-08-928-361B-1 Sequence 1, Appli
44 45.6 1.9 8920 2 US-08-446-855A-1 Sequence 1, Appli
45 45.6 1.9 8920 4 US-09-150-741-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-058-746-1/c
; Sequence 1, Application US/09058746
; Patent No. 6022716
; GENERAL INFORMATION:
; APPLICANT: Ilya Chumakov
; APPLICANT: Hiroaki Tanaka
; TITLE OF INVENTION: High Throughput DNA sequencing vector
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,746
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.015A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10317 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: CIRCULAR
; MOLECULE TYPE: synthetic DNA
; ORIGINAL SOURCE:
; ORGANISM: Cloning vector pGendel
; FEATURE:
; NAME/KEY: pGendel
; LOCATION: 1..10317
; FEATURE:
; NAME/KEY: Homology with X06404 compl (411..1668)
; LOCATION: 9..1266
; IDENTIFICATION METHOD: blastn against X06404
; FEATURE:
; NAME/KEY: Kanamycin resistance gene CDS

```

LOCATION: 142..957
IDENTIFICATION METHOD: By homology to X06404
FEATURE:
NAME/KEY: Tn1000's right end
LOCATION: complement 1332..1371
IDENTIFICATION METHOD: blastn against X60200)
FEATURE:
NAME/KEY: Homology with U46017 (1-472)
LOCATION: 1423..1894
IDENTIFICATION METHOD: blastn against U46017
FEATURE:
NAME/KEY: single stranded DNA replication origin
LOCATION: 1423..1894
IDENTIFICATION METHOD: By homology to U46017
OTHER INFORMATION: mutation T -> C 1658
FEATURE:
NAME/KEY: Homology with U51113 (2382..6997)
LOCATION: 1896..6544
IDENTIFICATION METHOD: blastn against U51113
FEATURE:
NAME/KEY: OriS
LOCATION: 1972..2188
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: repELR
LOCATION: 2897..2918
OTHER INFORMATION: Described in seqID 16
FEATURE:
NAME/KEY: RepE
LOCATION: 2903..3034
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: T3
LOCATION: 3043..3059
OTHER INFORMATION: Described in seqID 17
FEATURE:
NAME/KEY: LRT3RA
LOCATION: complement 3045..3069
OTHER INFORMATION: Described in seqID 15
FEATURE:
NAME/KEY: IncC
LOCATION: 3070..3320
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: Insertion 33 bases 3038..3071
FEATURE:
NAME/KEY: Para
LOCATION: 3655..4821
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: mutation G -> A 3878
FEATURE:
NAME/KEY: ParB
LOCATION: 4821..5792
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: ParC
LOCATION: 5865..6382
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: Homology with J01688 (complement 175..819)
LOCATION: 6574..7218
IDENTIFICATION METHOD: blastn against J01688
OTHER INFORMATION: mutation A -> G 7096
FEATURE:
NAME/KEY: CDS streptomycin sensitivity gene
LOCATION: complement 6716..7090
IDENTIFICATION METHOD: By homology to J01688
OTHER INFORMATION: mutation A -> G 6728
OTHER INFORMATION: mutation G -> C 6821
OTHER INFORMATION: mutation C -> T 6866
OTHER INFORMATION: mutation T -> C 7013
OTHER INFORMATION: mutation T -> A 7058
FEATURE:
NAME/KEY: rpsLR

```

```

LOCATION: 7155..7174
OTHER INFORMATION: Described in seqID 12
FEATURE:
NAME/KEY: SP6
LOCATION: 7230..7248
OTHER INFORMATION: Described in seqID 13
FEATURE:
NAME/KEY: Tn1000's left end
LOCATION: 7252..7291
IDENTIFICATION METHOD: blast (X60200)
FEATURE:
NAME/KEY: Homology with X02730 (complement 37..1959)
LOCATION: 7305..9227
IDENTIFICATION METHOD: blastn against X02730
FEATURE:
NAME/KEY: CDS levanucrase gene
LOCATION: complement 7379..8800
IDENTIFICATION METHOD: By homology to X02730
OTHER INFORMATION: mutation T -> C 7466
OTHER INFORMATION: mutation A -> G 7739
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347
OTHER INFORMATION: mutation T -> C 8600
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772
FEATURE:
NAME/KEY: SLR3
LOCATION: 8711..8731
OTHER INFORMATION: Described in seqID 14
FEATURE:
NAME/KEY: Homology with J01636 (complement 1158..1465)
LOCATION: 9298..9623
IDENTIFICATION METHOD: blastn against J01636
FEATURE:
NAME/KEY: CDS alpha peptide beta-galactosidase
LOCATION: complement 9276..9497
IDENTIFICATION METHOD: By homology to J01636
FEATURE:
NAME/KEY: primer HE1
LOCATION: complement 9465..9479
FEATURE:
NAME/KEY: primer HE2
LOCATION: 9461..9475
FEATURE:
NAME/KEY: primer LacRS2Avr
LOCATION: complement 9603..9630
FEATURE:
NAME/KEY: primer LacE2Mlu
LOCATION: 9289..9314
FEATURE:
NAME/KEY: Homology with M77789 (1889..2576)
LOCATION: 9629..10315
IDENTIFICATION METHOD: blastn against M77789
FEATURE:
NAME/KEY: high copy-number double-stranded DNA replication origin
LOCATION: complement 9629..10315
IDENTIFICATION METHOD: By homology to M77789
OTHER INFORMATION: mutation C -> T 9803
OTHER INFORMATION: site ScaI 10029 - 10034
OTHER INFORMATION: site PmlI 10038 - 10043
OTHER INFORMATION: CLONING SITES 10031 - 10041
FEATURE:
NAME/KEY: oriLRd
LOCATION: 9856..9881
OTHER INFORMATION: Described in seqID 8
FEATURE:
NAME/KEY: OS1
LOCATION: 10009..10026
OTHER INFORMATION: Described in seqID 10
FEATURE:
NAME/KEY: ORI
LOCATION: complement 10046..10062
OTHER INFORMATION: Described in seqID 11
FEATURE:
NAME/KEY: oriLR

```

LOCATION: complement 10182..10202
OTHER INFORMATION: Described in seqID 9
US-09-058-746-1

Query Match

Best Local Similarity 3.4%; Score 80; DB 3; Length 10317;
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

```

QY 983 cacaagaatgtcaggaatcagctgttttgaacagtgataactctatccaaattatttata 1042
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8322 CACAGAATGTGAGTTCAGCCACATTTACATCTGCGGAAAAATCCGTTTATTCTACA 8263
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1043 caagggtagacagctctgataaacaataccaatcatcaaaaaaattgtagcgactcttt 1102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8262 C-----TGATTTCTCCGGTAAACATTTACGCGCAACAAACACCTGCACAAAGTTA 8209
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1103 atttaactgataataatggaatgtatcatctcgtcaggtacgaaatgactatattgtat 1162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8208 ACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGPAGAGGATTTAAATCAATCT 8149
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1163 ttgaaggtgatggctattactacaaacttatgatacaatgga---aagctactaacaag 1219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8148 TTGACGGTGACGGAACAACTATCAAAATGTACAGAGTTCATCGATGAAGGCAACTACA 8089
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1220 gtccgataatatgcaatgcgtgatgctcatgtaattgaagatggtaatggtatcggt 1279
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8088 GCTCAGCGCAACACCATACGCTGAGAGATCCTCACTACGTAGATGAAGAGCCCAAAAT 8029
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1280 acctgttttgaagaagtactgtttggaataattcaagcgaggaacaaatttata 1339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8028 ACTTAGTATTGAGCAACACATGGAACCTGGAACCTGAAGATGGCTTACCAAGCGCAAGAACTTTAT 7969
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1340 actggttaattatggcgagatgacgcatattaatacaagagcttattgaattcttt 1399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7968 TTAACAAAGCATACTATGCAAAACACACATCATCTTCCGTCAGAAAGTCAAAAACCTC 7909
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1400 caaatgatgataatgaagtcgaggaacttgggctaagtcagctatcggtatcctcaaac 1459
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7908 TCGAAGCGATYAAAACGACGCGGTGATGTTAGCAACGGGCTCTCGGTATGATGAGC 7849
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1460 taataaggaagaaagaatcctcaagtgagcagaggtatactaccatttaattctgcac 1519
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7848 TAAACGATGATTACACTGAAAAAAGTGATGAA-----ACCGTGATTGCTCATCTA 7798
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1520 caatggtgaagcgaatgaagcgaacaaatgtagttaattaggttaataaataattact 1579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7797 ACACAGTACAGATGAATTTGAACGCGCAACGCTTTTAAATGAACGGCAATGGTATC 7738
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1580 tattgcccgtaccgcttttaaatcaggaagtaatatgatgctgttgatgaatgctaat 1639
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7737 TGTTCACTGACTCCCGCGGATCAAAATGACGATTGACGGCATTTACGCTCAACGATATT 7678
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1640 atgcccgttggtgataatgttgcgaatgctgagatgttgcgtgatgcttaactggtatt 1699
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7677 A-----CATGCTTGGTTATGTTTCTAATCTTTAACTTGGCCCAT 7639
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1700 ataagcgaatataatgactctgagtagtcttgcgtcttctgcctgcaaacctgacgga 1759
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7638 ACAAGCGCTGACAAACAACTGGCTTGTGTPTAAAAATGGATCTTTGATCCTCAACGATGTA 7579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1760 cagcaactttatcatattatgctgtcccggttgcgggaaagatgaccaagattagttta 1819
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7578 CTTTACTTACTACACTTCGCTGTACC---TCAACGGAAGGAACAAATGCGTGATTA 7522
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1820 ctctcatatgactaataaataagtagtagcgggtaaggaatggaattcaacttgggcac 1879
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7521 CAGCTATATGACAAACAGGAGATTCTACGCAACAAA-----CAATCAAGCTTGGC 7468
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1880 cgagtttcttactaacaataaaccgggataacacactactgttttagcttaaaatgacta 1939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7467 CGAGCTTCCTGCTGAACATCAAGGCAAGCAAAACATCTGTTGTCAAGACAGCATCTCTG 7408
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 1940 atcaaggggatt 1951
      ||||| ||||| |||||
Db 7407 AACAAAGGACAAAT 7396

RESULT 2
US-09-438-142-1/c
; Sequence 1, Application US/09438142
; Patent No. 6258571
; GENERAL INFORMATION:
; APPLICANT: Ilya Chumakov
; TITLE OF INVENTION: High Throughput DNA sequencing vector
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/438,142
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.015C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10317 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: synthetic DNA
; ORIGINAL SOURCE:
; ORGANISM: Cloning vector pGendel
; FEATURE:
; NAME/KEY: pGendel
; LOCATION: 1..10317
; FEATURE:
; NAME/KEY: Homology with X06404 compl (411..1668)
; LOCATION: 9..1266
; IDENTIFICATION METHOD: blastn against X06404
; FEATURE:
; NAME/KEY: Kanamycin resistance gene CDS
; LOCATION: 142..957
; IDENTIFICATION METHOD: By homology to X06404
; FEATURE:
; NAME/KEY: Tn1000's right end
; LOCATION: complement 1332..1371
; IDENTIFICATION METHOD: blastn against X60200)
; FEATURE:
; NAME/KEY: Homology with U46017 (1-472)
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: blastn against U46017
; FEATURE:
; NAME/KEY: single stranded DNA replication origin
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: By homology to U46017
; OTHER INFORMATION: mutation T -> C 1658
; FEATURE:
; NAME/KEY: Homology with U51113 (2382..6997)
; LOCATION: 1896..6544
; IDENTIFICATION METHOD: blastn against U51113

```

```

FEATURE:
NAME/KEY: OriS
LOCATION: 1972..2188
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: repELR
LOCATION: 2897..2918
OTHER INFORMATION: Described in seqID 16
FEATURE:
NAME/KEY: RepE
LOCATION: 2903..3034
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: T3
LOCATION: 3043..3059
OTHER INFORMATION: Described in seqID 17
FEATURE:
NAME/KEY: LRT3RA
LOCATION: complement 3045..3069
OTHER INFORMATION: Described in seqID 15
FEATURE:
NAME/KEY: IncC
LOCATION: 3070..3320
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: Insertion 33 bases 3038..3071
FEATURE:
NAME/KEY: Para
LOCATION: 3655..4821
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: mutation G -> A 3878
FEATURE:
NAME/KEY: ParB
LOCATION: 4821..5792
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: ParC
LOCATION: 5865..6382
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: Homology with J01688 (complement 175..819)
LOCATION: 6574..7218
IDENTIFICATION METHOD: blastn against J01688
OTHER INFORMATION: mutation A -> G 7096
FEATURE:
NAME/KEY: CDS streptomycin sensitivity gene
LOCATION: complement 6716..7090
IDENTIFICATION METHOD: By homology to J01688
OTHER INFORMATION: mutation A -> G 6728
OTHER INFORMATION: mutation G -> T 6866
OTHER INFORMATION: mutation C -> T 6821
OTHER INFORMATION: mutation T -> C 7013
OTHER INFORMATION: mutation T -> A 7058
FEATURE:
NAME/KEY: fpsLR
LOCATION: 7155..7174
OTHER INFORMATION: Described in seqID 12
FEATURE:
NAME/KEY: SP6
LOCATION: 7230..7248
OTHER INFORMATION: Described in seqID 13
FEATURE:
NAME/KEY: Tn1000's left end
LOCATION: 7252..7291
IDENTIFICATION METHOD: blast (X60200)
FEATURE:
NAME/KEY: Homology with X02730 (complement 37..1959)
LOCATION: 7305..9227
IDENTIFICATION METHOD: blastn against X02730
FEATURE:
NAME/KEY: CDS levansucrase gene
LOCATION: complement 7379..8600
IDENTIFICATION METHOD: By homology to X02730
OTHER INFORMATION: mutation T -> C 7466

```

```

OTHER INFORMATION: mutation A -> G 7739
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347
OTHER INFORMATION: mutation T -> C 8600
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772
FEATURE:
NAME/KEY: SLR3
LOCATION: 8711..8731
OTHER INFORMATION: Described in seqid 14
FEATURE:
NAME/KEY: Homology with J01636 (complement 1158..1465)
LOCATION: 9298..9623
IDENTIFICATION METHOD: blastn against J01636
FEATURE:
NAME/KEY: CDS alpha peptide beta-galactosidase
LOCATION: complement 9276..9497
IDENTIFICATION METHOD: By homology to J01636
FEATURE:
NAME/KEY: primer HE1
LOCATION: complement 9465..9479
FEATURE:
NAME/KEY: primer HE2
LOCATION: 9461..9475
FEATURE:
NAME/KEY: primer LacI RS2AVR
LOCATION: complement 9603..9630
FEATURE:
NAME/KEY: primer LacE2Mlu
LOCATION: 9289..9314
FEATURE:
NAME/KEY: Homology with M77789 (1889..2576)
LOCATION: 9629..10315
IDENTIFICATION METHOD: blastn against M77789
FEATURE:
NAME/KEY: high copy-number double-stranded DNA replication origin
LOCATION: complement 9629..10315
IDENTIFICATION METHOD: By homology to M77789
OTHER INFORMATION: mutation C -> T 9803
OTHER INFORMATION: site ScaI 10029 - 10034
OTHER INFORMATION: site PmlI 10038 - 10043
OTHER INFORMATION: CLONING SITES 10031 - 10041
FEATURE:
NAME/KEY: oriLRd
LOCATION: 9856..9881
OTHER INFORMATION: Described in seqid 8
FEATURE:
NAME/KEY: OS1
LOCATION: 10009..10026
OTHER INFORMATION: Described in seqid 10
FEATURE:
NAME/KEY: OR1
LOCATION: complement 10046..10062
OTHER INFORMATION: Described in seqid 11
FEATURE:
NAME/KEY: oriLRr
LOCATION: complement 10182..10202
OTHER INFORMATION: Described in seqid 9
US-09-438-142-1

```

	Query Match	3.4%	Score 80;	DB 4;	Length 10317;
	Best Local Similarity	47.7%;	Pred. No. 4e-11;		
	Matches 464;	Conservative	0;	Mismatches 460;	Indels 48; Gaps 6;
Qy	983	cacaagaatggtcaggatcagctgtttttgacagtgataaactctatcaaatatttata	1042		
Db	8322	CACAAGAATGTCAGGTTACGCCACATTTACATCTGACGGGAAAATCCGTTTATTCAC	8263		
Qy	1043	caagggttagacagctctgataaacaataccaatcatcaaaaaattgctagcgctactcttt	1102		
Db	8262	C-----TGATTTCTCCGGTAAACATTACGGCAACAAACACTGCACAACTGCACAAAGTTA	8209		
Qy	1103	atttaactgataaataaggaaatgtatcaactcgctcagggtacgaaatgactatatgtat	1162		

Db 8208 ACGTATACGATCAGACAGCTCTTTGAACATCAACGGTGTAGAGGATTATAAATCAATCT 8149
Qy 1163 ttgaaggtgagcttattactaccacacttatgatcaatgga---aagctactaacaag 1219
Db 8148 TTGACGGTGCAGGAAACAGTATGAAATGTACAGCAGTTTCATCGATGAAGGCAACTACA 8089
Qy 1220 gtgcgataatattgcaatgcgtgatctcgtatgaattgaatggaatggtgacggt 1279
Db 8088 GCTCAGGCGAACCATACGCTGAGAGATCTCTACTACGTAGAGATGATAAAGGCCACAAAT 8029
Qy 1280 acctgttttgaagaagtactggtttggaataattatcaagcgagggaccacaaattata 1339
Db 8028 ACTTAGTATTGAGCAACACACCTGGAACCTGAAGATGCTACCAAGCGGAGATCTTTAT 7969
Qy 1340 actggtttaattgagcgagatgagcatttaataatcaagagcttattagaattcttt 1399
Db 7968 TTAACAAAGCATACTATGGCAAAAGCACATCATCTTCGCTCAAGAAAGTCAAAAACATTC 7909
Qy 1400 ccaatgatgatataagagtcgagcaacttggtggaatgcagctatcggtatcctcaaac 1459
Db 7908 TGCAAAGCGATAAATAAACACGCGCTGAGTTAGCAAAACGCGCTCTCGGTATGATTGAGC 7849
Qy 1460 taataagagcagaagaatccttaagtgagagaggttatactcaccatttaattctgcac 1519
Db 7848 TAAACGATGATACACACTGAAAGAGTATGAA-----ACCGCTGATTGCATCTA 7798
Qy 1520 caatggaagcagataatgagcagcaaatgtagtataataggttaataataattact 1579
Db 7797 ACACAGTAACAGATGAAATGAAACGCGGAGCTCTTTAAATGAAACGGCAATGGTATC 7738
Qy 1580 tatttgcgctaccgttttaataatgaggaatgagtgatgctgttgatgaatgcttaatt 1639
Db 7737 TGTTCACTGACTCCCGCGATCAAAAATGACGATGACGGCATACGCTCAACGATATTT 7678
Qy 1640 atgcgctggtgataatgttgaatgttcgagatgttgcgagatgctgaatgagctactt 1699
Db 7677 A-----CATGCTGTGTTATGTTCTTAATCTTTTAACGTGGCCCAT 7639
Qy 1700 ataagcaattaaatgattctgagtagctgtgactgttctgttcttgcgaactggcgga 1759
Db 7638 ACAGCGCTGAACAAACTGGCTTGTGTTAAATATGGATCTTGATCTTAACGATGATA 7579
Qy 1760 cagcaacttattctattgtctcccggttgcggaagaatgacgaatgattgta 1819
Db 7578 CTTTACTTACTACACTTCGCTGTACC---TCAAGCGAAGGAAACAAATGTCGTGATTA 7522
Qy 1820 cttcatatgactaatagaaatggagtagcgggtgataaggaatggattcaacttggcgac 1879
Db 7521 CAAGCTATATGACAAACAGAGGATTCTACGCAGACAAA-----CAATCAACGTTTGGCG 7468
Qy 1880 cgaagttcttactacaaattacccggataacacaaactactgttttagctaaatgacta 1939
Db 7467 CGAGCTTCCTGCTGAACATCAAGGCAAGAAACATCTGTGTCAAAAGACAGCATCCTTG 7408
Qy 1940 atcaaggggatt 1951
Db 7407 AACAGGACAAAT 7396

RESULT 3

US-08-242-932-8
; Sequence 8, Application US/08242932
; Patent No. 5595740
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5595740-IgA Fc Binding Forms of
; the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL

COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,932
FILING DATE: 16-MAY-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF142
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3730 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-242-932-8

Query Match 2.9%; Score 68.8; DB 1; Length 3730;
Best Local Similarity 60.9%; Pred. No. 2.1e-08;
Matches 112; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 2065 aaacggcatcatcgtctacacaaatgatcctgaaacgcaactacaccagaaacccct 2124
Db 2609 ACACCGAGCGCTCCAGACACACCGCATGTTCCGGAATCACCAAGGCCCGCAGACCG 2668
QY 2125 gagacaccttaactccccaaacacacaaagactcctgaaatcctgggacacccctaaat 2184
Db 2669 CGTGTTCGGAATCACCAAGACTCCAGAGACCGCATGTTCCGGAATCACCAAGACT 2728
QY 2185 cctaatacactaactcggaaatccttttaactccgaaacgcttaagcaacctgaa 2244
Db 2729 CGAAGACGACCAAGATTCCGGAACCCCTTAAGACTCCAGACTCCTTAAGCTCCAGAC 2788
QY 2245 accc 2248
Db 2789 GTCC 2792

RESULT 4

US-08-714-481-8
; Sequence 8, Application US/08714481
; Patent No. 5766606
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5766606-IgA Fc Binding Forms of
; the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,481
FILING DATE: 16-SEP-1996

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-714-481-8

Query Match          2.9%; Score 68.8; DB 1; Length 3730;
Best Local Similarity 60.9%; Pred. No. 2.le-08;
Matches 112; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 2065 aaacgcgatgctctgctacacccaataatgctgaaacgccaactacacccagaacccct 2124
DB 2609 ACACCGCAGGCTCCAGACACACCGCATGTCGGGAATCACCNAAGGCCCCAGAGACCG 2668
QY 2125 gagacacctaatactcccaaaacacccaagaactcctgaaaatcctgggacacctcaact 2184
DB 2669 CGTGTTCGGGAATCACCNAAGACTCCAGAACGCCGATGTCGGGAATCACCNAAGACT 2728
QY 2185 cctaatacaccttaactctcggaattcttaactcagaacgcctaaagcaacctgaa 2244
DB 2729 CCAGAGACCAACCAAGATTCCGGACCCCTAAGACTCCAGACGTCCTTAAGCTTCCAGAC 2788
QY 2245 accc 2248
DB 2789 GTCC 2792

```

```

RESULT 5
PCT-US95-06111-8
; Sequence 8, Application PC/TUS9506111
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 186 Grinter Hall
; APPLICANT: City: Gainesville
; APPLICANT: State/Province: Florida
; APPLICANT: Country: US
; APPLICANT: Postal code/zip: 32611
; APPLICANT: Phone number: 904-392-8929
; APPLICANT: Fax number: 904-392-6600
; APPLICANT: Telex number:
; TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06111
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US95-06111-8

Query Match          2.9%; Score 68.8; DB 5; Length 3730;
Best Local Similarity 60.9%; Pred. No. 2.le-08;
Matches 112; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 2065 aaacgcgatgctctgctacacccaataatgctgaaacgccaactacacccagaacccct 2124
DB 2609 ACACCGCAGGCTCCAGACACACCGCATGTCGGGAATCACCNAAGGCCCCAGAGACCG 2668
QY 2125 gagacacctaatactcccaaaacacccaagaactcctgaaaatcctgggacacctcaact 2184
DB 2669 CGTGTTCGGGAATCACCNAAGACTCCAGAACGCCGATGTCGGGAATCACCNAAGACT 2728
QY 2185 cctaatacaccttaactcctcggaattcttaactcagaacgcctaaagcaacctgaa 2244
DB 2729 CCAGAGACCAACCAAGATTCCGGACCCCTAAGACTCCAGACGTCCTTAAGCTTCCAGAC 2788
QY 2245 accc 2248
DB 2789 GTCC 2792

```

```

RESULT 6
US-09-068-043-1
; Sequence 1, Application US/09068043
; Patent No. 6048694
; GENERAL INFORMATION:
; APPLICANT: MICHAEL GENE BRAMUCCI
; APPLICANT: VASANTHA NAGARAJAN
; TITLE OF INVENTION: POSITIVE SELECTION
; TITLE OF INVENTION: VECTOR FOR BACILLUS SP.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND
; ADDRESSEE: COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT OFFICE 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/068.043
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,201
; FILING DATE: NOVEMBER 3, 1995
; ATTORNEY/AGENT INFORMATION:

```


Best Local Similarity 53.9%; Pred. No. 3.1e-07;
Matches 132; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 2012 tacctggcgaaacgtgataaacactgttgattgggaacttaattggttattgattgattaaaaccgc 2071
||||| | | | | | | | | | | | | | | | | | | | |
Db 2405 TACCTCAAGCCCGCAGATACACCGAGGCTCCAGACACACCGCATGTTCGGGAATCACCAA 2464
||||| | | | | | | | | | | | | | | | | | | | |

QY 2072 atgatcctgtacacacaatgatcctgaaacgcccaactacacagaaacccctcgagacac 2131
||| | | | | | | | | | | | | | | | | | | | | | |
Db 2465 AGGCCCGAGAAGACCGCGTGTTCGGGAATCACAAAGACTCCAGAACCGCATGTTC 2524
||| | | | | | | | | | | | | | | | | | | | | | |

QY 2132 ctaatactcccacacacaaagactcctgaaatcctctgggacacctcaaatcctaata 2191
||| | | | | | | | | | | | | | | | | | | | | | |
Db 2525 CGGAATCACCAAGGCCCGAGAAGCACCGCGTGTTCGGGAATCACAAAGACTCCAGAA 2584
||| | | | | | | | | | | | | | | | | | | | | | |

QY 2192 cacttaatactcggaaatctcttaactccagaaacgccttaagcaacctgaaacccaaa 2251
||||| | | | | | | | | | | | | | | | | | | | |
Db 2585 CACCGCATGTTCGGGAATCACCAAGACTCCAGAACGACCAGATTCCGGAACCCCGTA 2644
||| | | | | | | | | | | | | | | | | | | | | | |

QY 2252 ctaat 2256
||| | | | | | | | | | | | | | | | | | | | | | |
Db 2645 AGACT 2649
||| | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-08-923-992A-9
; Sequence 9, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 03-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3492 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3492
US-08-923-992A-9

Query Match 2.7%; Score 64.2; DB 1; Length 4200;
Best Local Similarity 53.9%; Pred. No. 3.4e-07;
Matches 132; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 2012 tacctggcgaaacgtgataaacactgttgattgggaacttaattggttattgattgattaaaaccgc 2071
||||| | | | | | | | | | | | | | | | | | | | |
Db 2513 TACCTCAAGCCCGCAGATACACCGAGGCTCCAGACACACCGCATGTTCGGGAATCACCAA 2572
||||| | | | | | | | | | | | | | | | | | | | |

QY 2072 atgatcctgtacacacaatgatcctgaaacgcccaactacacagaaacccctcgagacac 2131
||| | | | | | | | | | | | | | | | | | | | | | |
Db 2573 AGGCCCGAGAAGACCGCGTGTTCGGGAATCACAAAGACTCCAGAACCGCATGTTC 2632
||| | | | | | | | | | | | | | | | | | | | | | |

QY 2132 ctaatactcccacacacaaagactcctgaaatcctctgggacacctcaaatcctaata 2191
||| | | | | | | | | | | | | | | | | | | | | | |
Db 2633 CCCAATCACCAAGGCCCGCAGAACACCGCGTGTTCGGGAATCACCAAGACTCCAGAA 2692
||||| | | | | | | | | | | | | | | | | | | | |

QY 2192 cacttaatactcggaaatctcttaactccagaaacgccttaagcaacctgaaacccaaa 2251
||||| | | | | | | | | | | | | | | | | | | | |
Db 2693 CACCGCATGTTCGGGAATCACCAAGACTCCAGAACGACCAGATTCCGGAACCCCGTA 2752
||| | | | | | | | | | | | | | | | | | | | | | |

QY 2252 ctaat 2256
||| | | | | | | | | | | | | | | | | | | | | | |
Db 2753 AGACT 2757
||| | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
US-08-242-932-1
; Sequence 1, Application US/08242932
; Patent No. 5595740
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5595740-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/242,932
; FILING DATE: 16-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI42
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-242-932-1

Query Match 2.7%; Score 64.2; DB 1; Length 4200;
Best Local Similarity 53.9%; Pred. No. 3.4e-07;
Matches 132; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 2012 tacctggcgaaacgtgataaacactgttgattgggaacttaattggttattgattgattaaaaccgc 2071
||||| | | | | | | | | | | | | | | | | | | | |
Db 2513 TACCTCAAGCCCGCAGATACACCGAGGCTCCAGACACACCGCATGTTCGGGAATCACCAA 2572
||||| | | | | | | | | | | | | | | | | | | | |

Db	2832	TACCTCAAGCCCGGATACACCGCAGGTCCAGACACACCGCATGTTCCGGAA	2891
Qy	2072	atgatcctgtacacccaatgatctgaacgccaactataccagaaacccctgagac	2131
Db	2892	AGCCCCAAGACGCGGTGTTCCGGAATCACCAAAGACTCCAGAAAGCACCGCATGTC	2951
Qy	2132	ctaatactcccaaacacccaagactcctgaaatcctctgggacactcgaactcctaata	2191
Db	2952	CGGAATCACCAAAGGCCCCAGAACCGCGGTGTTCCGGAATCACCAAAGACTCCAGAG	3011
Qy	2192	cacctaaactccggaattcctttaactcagaaacgcctaagcaacctgaaacccaaa	2251
Db	3012	CACCCGATGTTCCGGAATCACCAAAGACTCCAGAAAGCACCAAGATTC	3071
Qy	2252	ctaata	2256
Db	3072	AGACT	3076

```

RESULT 12
US-08-714-481-1
; Sequence 1, Application US/08714481
; Patent No. 5766606
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5766606-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,481
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-714-481-1

```

		Query Match	2.7%;	Score 64.2;	DB 1;	Length 4200;
		Best Local Similarity	53.9%;	Pred. No. 3.4e-07;		
		Matches 132;	Conservative	0;	Mismatches 113;	Indels 0;
						Gaps 0;
Qy	2012	tacctggcggaactgataaacctgtgtattggatggaattggtatggattataaacccgc	2071			
Db	2832	TACCTCAAGCCCGAGATACACCGCAGGCTCCAGACACACCGCATGTTCCGGAAATCACCAA	2891			
Qy	2072	atgatcctgtacaccaaattgatccttgaacgcgcacactacacagaataaccccttgagac	2131			

Db	2892	AGGCCCCAGAAAGCACC	CGGTTC	CGGAATC	ACCAAAGACT	CCAGAAGC	ACCGCATGTC	2951
Qy	2132	ctaatactcccaaaacac	caaaagactcct	gaaaaatcct	ggagacac	ctcaaaactc	taata	2191
Db	2952	CGGAATC	ACCAAAGGCC	CCAGAAGC	ACCGCGT	GTTC	CGGAATC	ACCAAAGACT
Qy	2192	cacctaatctccggaatt	ccttttaactcc	agaagcgc	ctaaagca	acctggaac	cccaaa	2251
Db	3012	CACCGATG	TTC	CGGAATC	ACCAAAGACT	TCCAGAAGC	ACCAAAGATT	CCGGAACCCCTA
Qy	2252	ctaat	2256					3071
Db	3072	AGACT	3076					

RESULT 13

PCT-US95-06111-1

; Sequence 1, Application PC/TUS9506111

; GENERAL INFORMATION:

; APPLICANT:

; APPLICANT: Street address: 186 Grinter Hall

; APPLICANT: City: Gainesville

```

RESULT 13
PCT-US95-06111-1
; Sequence 1, Application PC/TUS9506111
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 186 Grinter Hall
; APPLICANT: City: Gainesville
; APPLICANT: State/Province: Florida
; APPLICANT: Country: US
; APPLICANT: Postal code/Zip: 32611
; APPLICANT: Phone number: 904-392-8929
; APPLICANT: Fax number: 904-392-6600
; APPLICANT: Telex number:
; TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06111
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-06111-1

```

Query Match 2.7% Score 64.2; DB 5; Length 4200;
Best Local Similarity 53.9%; Pred. No. 3.4e-07;
Matches 132; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
Qy 2012 tacctggcgaaactgtataaaccttttattgggacttaattggtatggattaaaaccgc 2071

```

Db 2832 TACCTCAAGCCCGAGATACACCGAGGCTCCAGACACACCGCATGTTCCGGAATCACCAA 2891
QY 2072 atgactcgtctacacaaatgatctctgaaagccaaactacacagaaacccctgagacac 2131
Db 2892 AGGCCCGAGACACCGCGTGTTCGGAATCACCAAGACTCCAGAGACCGCATGTTCC 2951
QY 2132 ctaatactcccaaaacacaaagactcctgaaatcctggaacacccctaaactcctaata 2191
Db 2952 CGGAATCACCAAGGCCCGAGACACCGCGTGTTCGGAATCACCAAGACTCCAGAG 3011
QY 2192 cactaatactcgggaattccttaactccagaaacgccttaagcaaacctgaaacccaaa 2251
Db 3012 CACCGATGTTCCGGAATCACCAAGACTCCAGAGACCGCATGTTCCGGAATCACCA 3071
QY 2252 ctaat 2256
Db 3072 AGACT 3076

```

RESULT 14

```

US-08-923-992A-1
; Sequence 1, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923.992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 320..3811
US-08-923-992A-1

```

```

Query Match . 2.7%; Score 63.8; DB 4; Length 4200;
Best Local Similarity 53.9%; Pred. No. 4.3e-07;
Matches 131; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```

```

QY 2006 ctgctttacctggcgaaacgtgataaacctgttgattgggacttaattggtatggattaa 2065
Db 2880 CGGAATCACCAAGGCCCGAGAGACACCGCGTGTTCGGAATCACCAAGACTCCAGAG 2939
QY 2066 aaccgcatgatcctgctacacccaaatgatcctgaaacgcgaacacacacccctg 2125
Db 2940 CACCGCATGTTCCGGAATCACCAAGGCCCGAGAGACACCGCGTGTTCGGAATCACCA 2999
QY 2126 agaaccttaatactcccaaaacacacacacacacacacacacacacacacacacacac 2185
Db 3000 AGACTCCAGAGACACCGCATGTTCCGGAATCACCAAGACTCCAGAGACCGCATGTTCC 3059
QY 2186 ctaatacactaatactcgggaattccttaactccagaaacgccttaagcaaacctgaaa 2245
Db 3060 CGAAACCCCTTAAGACTCCAGAGACCGTCCCTAAGCTTCAGAGCTTCCTAAGCTTCAGAG 3119
QY 2246 ccc 2248
Db 3120 TCC 3122

```

RESULT 15

```

US-07-638-431-1
; Sequence 1, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumthi, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: Immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:
; LIBRARY: Py-lambda gtl1-2-7 kb genomic expression

```

; CLONE: Py10.1111
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 718..3195
; OTHER INFORMATION:
US-07-638-431-1

Query Match 2.7%; Score 63.2; DB 1; Length 4673;
Best Local Similarity 54.2%; Pred. No. 6.5e-07;
Matches 128; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
Qy 2061 attaaaaccgcatgctctgtacacccaaatgctcctgaaacgccaactacacccagaaac 2120
Db 1626 AGTAAATCCAAACGATCCAAACGATCCAAACACCCCAACCCCAATACCCCAACCAA 1685
Qy 2121 ccctgagacaccttaatactcccaaaacacaaagactcctgaaaatcctgggacacctca 2180
Db 1686 CCCAAACACCCCAATACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAAC 1745
Qy 2181 aactcctataacacctaatactccggaattccttaactccagaaacgccaagcaacc 2240
Db 1746 CAATCCAAATACCCCAATACCCCAACCCCAATACCCCAATACCCCAATACCCCAAC 1805
Qy 2241 tgaaccccaactaataatcggttgccacaaactggaataatgccaataagcca 2296
Db 1806 AATAACCCCAACACCCCAATACCCCAATACCCCAATACCCCAATACCCCAATACCCCA 1861

Search completed: September 22, 2002, 05:22:06
Job time: 2066Y sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 01:51:46 ; Search time 1031.3 Seconds
(without alignments)
3940.596 Million cell updates/sec

Title: US-09-995-587A-2

Perfect score: 2367

Sequence: 1 atgtataaaagcggtaaaaa.....ttaacaaacgtcgatttaac 2367

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	3.4	10317	21	AAZ35988
2	78.4	3.3	1668	21	AAZ57329
3	78.4	3.3	1722	21	AAZ57330
4	71.8	3.0	2537	23	AAZ82323
5	71.8	3.0	2787	23	AAZ77631
6	71.8	3.0	2787	23	AAZ86008
7	71.8	3.0	2787	23	AAZ86581
8	71.8	3.0	2787	23	AAZ88736
9	71.8	3.0	2787	23	AAZ91108
					SacB and cpy chime
					SacB and cpy chime
					DNA encoding novel
					DNA encoding novel
					DNA encoding novel
					DNA encoding novel
					DNA encoding novel
					DNA encoding novel

10	70.2	3.0	2079	23	AAZ585786	DNA encoding novel
11	70.2	3.0	2079	23	AAZ58598	DNA encoding novel
12	70.2	3.0	2079	23	AAZ87342	DNA encoding novel
13	70.2	3.0	2079	23	AAZ88703	DNA encoding novel
14	69	2.9	2325	23	AAZ89777	DNA encoding novel
15	69	2.9	2325	23	AAZ75941	DNA encoding novel
16	69	2.9	3084	23	AAZ85832	DNA encoding novel
17	69	2.9	3084	23	AAZ87383	DNA encoding novel
18	68.8	2.9	3730	17	AAZ03195	DNA encoding novel
19	68.2	2.9	2745	23	AAZ73811	Group B Streptococ
20	68.2	2.9	5912	23	AAZ86048	DNA encoding novel
21	67.8	2.9	1690	23	AAZ85792	DNA encoding novel
22	67.8	2.9	1690	23	AAZ86019	DNA encoding novel
23	67.8	2.9	3288	23	AAZ87350	DNA encoding novel
24	67.8	2.9	3288	23	AAZ77395	DNA encoding novel
25	67.8	2.9	3288	23	AAZ85789	DNA encoding novel
26	67.8	2.9	3288	23	AAZ86010	DNA encoding novel
27	67.8	2.9	3288	23	AAZ87346	DNA encoding novel
28	67.8	2.9	3288	23	AAZ88746	DNA encoding novel
29	67.4	2.8	2588	23	AAZ86887	DNA encoding novel
30	67.4	2.8	3305	18	AAZ68806	Mutant levansucras
31	67.4	2.8	6721	24	AAZ18599	Purinergic recepto
32	67.4	2.8	6721	24	AAZ18600	Purinergic recepto
33	67.2	2.8	1584	23	AAZ68494	DNA encoding novel
34	67.2	2.8	1629	23	AAZ85800	DNA encoding novel
35	67.2	2.8	1629	23	AAZ86022	DNA encoding novel
36	67.2	2.8	1629	23	AAZ87359	DNA encoding novel
37	67.2	2.8	1935	23	AAZ70417	DNA encoding novel
38	67.2	2.8	1974	23	AAZ85799	DNA encoding novel
39	67.2	2.8	1974	23	AAZ86021	DNA encoding novel
40	67.2	2.8	2044	23	AAZ85834	DNA encoding novel
41	67.2	2.8	2044	23	AAZ87384	DNA encoding novel
42	67.2	2.8	2328	23	AAZ86428	DNA encoding novel
43	67.2	2.8	2361	23	AAZ66444	DNA encoding novel
44	67.2	2.8	2361	23	AAZ85823	DNA encoding novel
45	67.2	2.8	2370	23	AAZ69133	DNA encoding novel

ALIGNMENTS

RESULT 1
AAZ35988/c
ID AAZ35988 standard; DNA; 10317 BP.
XX
AC AAZ35988;
XX
DT 09-FEB-2000 (first entry)
XX
DE Vector pGendell sequence SEQ ID NO:1.
XX
KW DNA sequencing vector; nested deletion; transposition; contig;
KW truncated lacZ; origin of replication; pGendell; mapping; ss.
XX
OS Synthetic.
XX
PN WO9953044-A2.
XX
PD 21-OCT-1999.
XX
PF 09-APR-1999; 99WO-IB00690.
XX
PR 10-APR-1998; 98US-0058746.
XX
PA (GEST) GENSET.
XX
PI Chumakov I, Tanaka H;
XX
DR WPI; 2000-023040/02.
XX
PT New DNA sequencing vectors, used for sequencing large regions of DNA
XX and mapping the location of markers

Claim 43; Page 93-96; 102pp; English.

Claim 43: Page 93-96; 102pp; English.

The present invention describes a vector comprising the following:

(a) a high copy number origin of replication (Ori) having at least one cloning site, with at least one cloning site being positioned in the high copy number Ori such that the ability of the high copy number Ori to direct replication is not disrupted when no insert has been cloned into the cloning site and is disrupted when an insert is cloned into the cloning site; and (b) a low copy number Ori. Also described is a truncated lacZ gene to distinguish cells with the truncated lacZ gene at high copy number from cells carrying it at a low copy number, where the former are dark blue when grown on medium containing xgal and IPTG, and the latter are light blue when grown on medium containing xgal and IPTG. Products and methods from the present invention can be used for the manipulation of DNA. They can be used for sequencing large regions of DNA and mapping the locations of markers within large regions of DNA. The present sequence represent a specifically claimed vector sequence from the present invention.

Sequence 10317 BP; 2567 A; 2296 C; 2469 G; 2985 T; 0 other;

Query Match 3.48; Score 80; DB 21; Length 10317:

Best Local Similarity 47.7%; pred. No. 2.5e-08;

Matches	464	Conservative	0	Mismatches	460	Indels	48	Gaps	6
---------	-----	--------------	---	------------	-----	--------	----	------	---

QY	983	cacaaagaatggtcaggaatcagctgttttgaacagtgataactctatccaatattttata	1042
Db	8322	CACAAGAATGGTCAGGTTTCAGCCACATTACATCTGACGGAAAAATCCGTTTATTCTACA	8263
QY	1043	caagggtagacgcgtctgataaacaataccaatcatcaaaaaattgctagcgtactctttt	1102
Db	8262	C-----TGATTTCTCCGTTAAACATTTACGCCAACAACAACCTGACAACTTGCACAGTTA	8209
QY	1103	atttaactgataataatggaatgtatcactcgcctcaggtcacgaaatgactatattgtat	1162
Db	8208	ACGTATCAGCATCAGACAGCTCTTTTGACATCAACGGGTGAGAGGATTATAAATCAATCT	8149
QY	1163	ttgaagggtgatggtctatTactaccaaaacttatgatcaatgga---aagctactaacaaag	1219
Db	8148	TTGACGGTGACGGAAAAACGTATCANAATGTACAGAGTTCATCGATGAGGCCAATACA	8089
QY	1220	gtgcgataattatgcaatgcgtgatctcgaatgtaattgaagatggtaattggatcggt	1279
Db	8088	GCTCAGGGGACAAACATACGCTGAGAGATCCTCACTACGTAGAAGATAAAAGGCCACAAAT	8029
QY	1280	acctgtttttgaagcaagtaactggtttgaaaaattatcaagcgaggaaccaaaattata	1339
Db	8028	ACTTAGTATTTGAAGCAACACTGGAACTGAAGTGGCTACCAAGGCGAAGAATCTTTAT	7969
QY	1340	actggttaaatattgoggaagatgacgcattaatatcaagagcttatttagaattctttt	1399
Db	7968	TTAACAAAGACATATTATGGCAAAAAGCACATCTTCCGTCGAAGAAAGTCAAAAACCTTC	7909
QY	1400	ccaatgatgatattaagagtcgcggaacttggctaatgcagctatcggtatcctcaaac	1459
Db	7908	TGCAAAAGCGATAAAAACGCCAGCGCTCAGTTACGAACGGCGCTCTCGGTATGATTGAGC	7849
QY	1460	taataaaggacgaaagaatcctaaggtggcagagttatactcaccattaaattctcgac	1519
Db	7848	TAAACGATGATTACACACTGAAAAAAGATGATGAA-----ACCGCTGATTGCATCTA	7798
QY	1520	caatggtaagcgatgaattgaugcaccaaaatagttaaattaggtaataataattact	1579
Db	7797	ACACAGTAAACAGATGAATTTGAACCGCGAACGCTCTTAAATGAACGGCAATGGTGATC	7738
QY	1580	tatttgcgcgtaccctgtttaaactcgagggaagtgaatgatgcttggatgaatgctaaat	1639
Db	7737	TGTTCACTGACTCCCGGGATCAAAAATGACGATTGACGGCATTAGCTCTACGATATTT	7678
QY	1640	atgcgcgttgggtgaatagttgcgaatggcggatattgttgcgtatagatcctaactggatctt	1699
Db	7677	A-----CATGCTGGTTGATCTTTCTAACTCTTAACTCTTTAACTGGCCCAT	7639

Query Match 3.3%; Score 78.4; DB 21; Length 1668;
Best Local Similarity 47.6%; Pred. No. 3.6e-08;
Matches 463; Conservative 0; Mismatches 461; Indels 48; Gaps 6;

QY 983 cacaagaatggtcaggatcagctgttttgaacagtgatgaactctctcaatattttata 1042
|||||
Db 725 cacaagaatggtcaggatcagctgttttgaacagtgatgaactctctcaatattttata 784
|||||

QY 1043 caagggtagacagctgtgataaacaataccaatccaataatgtagcgtactcttt 1102
|||||
Db 785 c-----tgattcttcggttaaacattacggtgcaaacacacgacagcaagta 838
|||||

QY 1103 atttaactgataaataaggaaatgatacactcgtcaggtacgaaatgactatattgtat 1162
|||||
Db 839 acgtatcagcatcagacagctctttgaacatcaacggtgtagaggattataaatcaatct 898
|||||

QY 1163 ttgaaggtgagctattactaccacacttatgatcaatgga---aagctactaacaag 1219
|||||
Db 899 ttgacgtgacggaacacgttacaaatgtacagcagttcatcgatgaagcgaactaca 958
|||||

QY 1220 gtccgataatattgaatcgctgagctcgtcatgtaattgaagtggtatgctggt 1279
|||||
Db 959 gctcagcgacacaccatcagctgagagatcctcactacgttagaagataaaggccacaat 1018
|||||

QY 1280 acctgttttgaagcaagctactggtttgaaattatcaagcgagggaccacaaattata 1339
|||||
Db 1019 acttagtatttgaagcaaacactcggaaactggaagatggctaccaggcggaagatctttat 1078
|||||

QY 1340 actggttaattatggtggagatgagcagcatttaatacaagagcttatttagaattcttt 1399
|||||
Db 1079 ttaacaaagcactactggaagacacatcattcttcgtcgaagaaagtcacaaacttc 1138
|||||

QY 1400 ccaatgatgataataagtcggtggcaacttgggttaagtcagctatcgttactcctcaaac 1459
|||||
Db 1139 tgcaagcgataaaacacgcgcgtgagttagtaacacggtcgtcgtgtagtgagc 1198
|||||

QY 1460 taaataagcagcaaaagaaatccttaagtggtgagagttatactcaccatattttctgcac 1519
|||||
Db 1199 taaacagatattacacactgaaagagtgatgaa-----acgcgtgagtgcatcta 1249
|||||

QY 1520 caatggtgaagcagaaatggcgccacaaatgtagttaaattaggtataataattact 1579
|||||
Db 1250 acacagtaacagatgaattgaacgcgcgaacgtctttaaataagacgcaaatggtacc 1309
|||||

QY 1580 tatttgcctaccctgttaaaatcgaggaagtaagatgatgctgttgatgaaatgctaat 1639
|||||
Db 1310 tgttcaactgactcccgcgatcaaaaatgacgattgacggttactcgttaacgatat 1369
|||||

QY 1640 atgcggttgatgaatgttgcaatggtcggatattgttgatgattgctaaactgattct 1699
|||||
Db 1370 a-----catgctggttatgttttctaatcttcaactggtcccat 1408
|||||

QY 1700 ataagcattaaatgattcgtgagtagtcttctgactgtcttctgctgcacaaactggcga 1759
|||||
Db 1409 acaagcgtgacaaacactgctgtgtttaaataaggtactcttgatcctcaacgagtaaa 1468
|||||

QY 1760 cagcaacttattcattatgctgtcccggttgcggaaagaaagatgaccagattattgta 1819
|||||
Db 1469 ccttacttactcaacttcgctgtacc---tcaagcgaagaaagaaatgctgtgattta 1525
|||||

QY 1820 cttcatatgactaatagaatggagtagcgggttaaaggaatggtattcaacttgggcac 1879
|||||
Db 1526 caagctatatgacaacagagattctacgagacaaa-----caatcaacgtttgctgc 1579
|||||

QY 1880 cgaagtcttactcaaaataacccggataacacactactgttttttagtcaaaatgacta 1939
|||||
Db 1580 caagcttctgctgtaacatcaaggcaagaaacacatctgttgtcaaaagacagcatccttg 1639
|||||

QY 1940 atcaaggggatt 1951
|||||

Db 1640 aacaaggacaat 1651
|||||

RESULT 3
AAZ57330
ID AAZ57330 standard; DNA; 1722 BP.
XX AC AAZ57330;
XX DT 03-APR-2000 (first entry)
XX DE SacB and cpy chimeric gene #2.
XX KW Levansucrase; SacB; cpy; salt tolerance; vacuole guide peptide;
KW carboxypeptidase A; chimeric gene; transgenic plant; yeast;
KW Bacillus subtilis 168; Saccharomyces cerevisiae X8;
XX drought resistance; ds.
XX Chimeric - Saccharomyces cerevisiae.
OS Chimeric - Bacillus subtilis.
XX CN1231337-A.
XX 13-OCT-1999.
XX 08-APR-1998; 98CN-0101336.
XX 08-APR-1998; 98CN-0101336.
XX (GENE-) INST GENETICS CHINESE ACAD SCI.
XX Chen S, Zhang H, Dong W;
XX WPI; 2000-087902/08.
XX Method for improving plant salt resistance using transgenic technology
XX Example; Fig 4; 12pp; Chinese.
XX A method has been developed for raising salt tolerance of plants using
transgenic technology. The method comprises: cloning levansucrase gene
SacB and vacuole guide peptide (cpy) sequence of carboxypeptidase A
from Bacillus and DNA of yeast, using the two kinds of gene to create a
chimeric gene, and using the chimeric gene to structure plant expression
carrier, using the obtained plant expression carrier to transform a
plant and screen for resistant seedlings. The method can obtain
drought-resistant salt-tolerance plants, and the polymerase chain
reaction (PCR) and Northern analysis of transgenic plants shows that
the exogenous gene is integrated in the salt-tolerance transgenic plant.
The present sequence represents a SacB and cpy chimeric gene from the
present invention.
XX Sequence 1722 BP; 590 A; 390 C; 339 G; 403 T; 0 other;
SQ

Query Match 3.3%; Score 78.4; DB 21; Length 1722;
Best Local Similarity 47.6%; Pred. No. 3.6e-08;
Matches 463; Conservative 0; Mismatches 461; Indels 48; Gaps 6;

QY 983 cacaagaatggtcaggatcagctgttttgaacagtgatgaactctctcaatattttata 1042
|||||
Db 756 cacaagaatggtcaggatcagctgttttgaacagtgatgaactctctcaatattttata 815
|||||

QY 1043 caagggttagacagctgtgataaacaataccaatccaataatgtagcgtactcttt 1102
|||||
Db 816 c-----tgattcttcggttaaacattacggtgcaaacacacgacagta 869
|||||

QY 1103 atttaactgataaataaggaaatgatacactcgtcaggtacgaaatgactatattgtat 1162
|||||
Db 870 acgtatcagcatcagacagctctttgaacatcaacggtgtagaggattataaatcaatct 929
|||||

QY 1163 ttgaaggtgagctattactaccacacttatgatcaatgga---aagctactaacaag 1219
|||||
Db 930 ttgacgtgacggaacacgttacaaatgtacagcagttcatcgatgaagggcaactaca 989
|||||

QY 1220 gtccgataattgcaatcggtgatgctcatgtaattgaagatggtgaatggtgatcgg 1279
 Db 990 gctcaggcgacaccatacgtgagatcctcactacgtagaagataaaggccacacaaat 1049
 QY 1280 acctgttttgaagcaagctactggtttggaattatcaaggcgagacacaaatttata 1339
 Db 1050 acttagtattgagcaaacactggaactgaagtgaagctaccacgaagcggaagatctttat 1109
 QY 1340 actggttaattatggtgagagacgcatttaataatcaagagccttatttagaattcttt 1399
 Db 1110 ttacaagaacatactatgcaaaagcacatcttctccgtcaagaagtcacaaacttc 1169
 QY 1400 ccaatgatgattatgaagtcggcgcaacttgggtcctaagtcagctatcggtatctctcaaac 1459
 Db 1170 tgcaagcgataaaacacgcacgctgagtgtagcaaacggtcctcgtgatgttgagc 1229
 QY 1460 taataaggcgaagaaatcctaaggtggcagagttatactcacacattatttctgcac 1519
 Db 1230 taacgatgattacacactgaaaaagtgtgaa-----accgtgattgcatcta 1280
 QY 1520 caatgtaagcgatgaattgacgaccaaataatgattgtaattaggttaataataattact 1579
 Db 1281 acacagtaacagatgaattgaaacgcgcgaactctttaaataagcggcaaatggtacc 1340
 QY 1580 tattgcgctaccctgttaaatcgaggaagtaataatgatgatgctgttgatgaatgctaatt 1639
 Db 1341 tgttactgactcccggaatcaaaaatgacgattgacgacattacgtcttaacgatat 1400
 QY 1640 atgctgttggtgataattgttgcgaatggtcgatgattgctgtagtcttaactggtatctt 1699
 Db 1401 a-----catgctgttattgtttcttaattcttaactggtcccat 1439
 QY 1700 ataagccataaatgattctggagtagtctgactgcttctgtcttcgcacaaatggcgga 1759
 Db 1440 acaagcgtgacaaacactggcctgtgtgtaaaatggatcttgcacccaacgatgtaa 1499
 QY 1760 cagcaactattcatattatgctgtccctgttgcggaaaaagatgacaaagtattagta 1819
 Db 1500 ccttactactcacactcgtcgtacc-----tcaaggcaaggaaacaaatgctgatta 1556
 QY 1820 cttcatatgactaataagaatggagtagcggtgtaaaaggaaatggattcaacttggcgac 1879
 Db 1557 caagctatgacaaacagaggattctacgcagacaaa-----caatcaactgttgcgc 1610
 QY 1880 cgaatttcttactacaaattacccggataacacaaactactgttttagctaaatgacta 1939
 Db 1611 caagcttctcgtgaacatcaaaaggcaaaaacatctgtgtcaaaagacagcatccttg 1670
 QY 1940 atcaaggggatt 1951
 Db 1671 aacaaggacaaat 1682

RESULT 4

AAS82323
 ID AAS82323 standard; cDNA; 2537 BP.

XX AC AAS82323;

XX AC AAS82323;

DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #18127.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX XX WO200175067-A2.

XX PN 11-OCT-2001.

XX PD 30-MAR-2001; 2001WO-US08631.

XX PF

XX

PR

PR

XX

PA

XX

XX

PI

XX

DR

DR

PT

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

31-MAR-2000; 2000US-0540217.
 23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

P-PSDB; ABG18136.

New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity -

Claim 1; SEQ ID No 18127; 103pp; English.

The invention relates to isolated polynucleotide (I) and
 polypeptide (II) sequences. (I) is useful as hybridisation probes,
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 and gene mapping, and in recombinant production of (II). The
 polynucleotides are also used in diagnostics as expressed sequence tags
 for identifying expressed genes. (I) is useful in gene therapy techniques
 to restore normal activity of (II) or to treat disease states involving
 (II). (II) is useful for generating antibodies against it, detecting or
 quantitating a polypeptide in tissue, as molecular weight markers and as
 a food supplement. (II) and its binding partners are useful in medical
 imaging of sites expressing (II). (I) and (II) are useful for treating
 disorders involving aberrant protein expression or biological activity.
 The polypeptide and polynucleotide sequences have applications in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits to assess biodiversity
 and to produce other types of data and products dependent on DNA and
 amino acid sequences. AAS64197-AAS94564 represent novel human
 diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

Sequence 2537 BP; 681 A; 608 C; 707 G; 541 T; 0 other;

Query Match

Best Local Similarity 3.0%; Score 71.8; DB 23; Length 2537;
 Matches 453; Conservative 0; Mismatches 482; Indels 42; Gaps 5;

QY 983 cacaagaatggtcagatcagctgtttgaaacagtgatacctctatcccaattttata 1042
 Db 1069 cgcaagaatggtcgggttctgcaacctttacatctgacggaaaaatccgtttatttaca 1128
 QY 1043 caagggtagacacgctctgataaacaataccaatacaaaaaattgctagcgtactcttt 1102
 Db 1129 ctgactattccggttaaacattacggaacaaagcctgacacagcgagggttaattgt 1188
 QY 1103 atttaactgataaataatggaaatgtatcactcgtcaggtacgaaatgactatctgtat 1162
 Db 1189 caaaatctgatgacacactcaaaaatacaacggagtggaagatcacaaaacgattttgacg 1248
 QY 1163 ttgaaggtgatggtctattactaccacaaacttatgcaatggaagctactacaagaagtg 1222
 Db 1249 gagcggaaaaacataatcagaacgtttcattttttttatcgatgaaggcaattatcacatcg 1308
 QY 1223 ccgataaattgcaatcggtgatgctcgtatgtaattgaagatggttaattgctcggtacc 1282
 Db 1309 ggcacaacacacgctgagagacccctcactacgtttgaga---caaggccataataacc 1365
 QY 1283 ttgtttttgaagcagagtagtactggtttgaaaaattatcaaggcagaccacaaattataact 1342
 Db 1366 ttgtattcgaaagccaaacacgggaacagaaacggataccaaagcggaagatcttttatta 1425
 QY 1343 ggttaaatattggcgagatgacgcatttaataatcaagagcttatttagaattctttcca 1402

Db 1283 aaggacagctgacatgg 1299

RESULT 9
 AAS91108
 ID AAS91108 standard; cDNA; 2787 BP.

XX AAS91108;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #26912.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PSDB; ABG26921.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity

PS Claim 1; SEQ ID NO 26912; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2787 BP; 935 A; 596 C; 607 G; 649 T; 0 other;

Query Match
 Best Local Similarity 3.0%; Score 71.6; DB 23; Length 2787;
 Matches 453; Conservative 0; Mismatches 482; Indels 42; Gaps 5;

Qy 983 cacaagaatggtcaggatcagctgttttgacagtgataactctatccaattatttata 1042

Db 365 cgcagaatggtcccggttctgcaacctttacatgtgcggaataatccgtttattctaca 424

Qy 1043 caagggtagacacgtctgtgataacaataccaatcatcaaaaaattgtctagcgctactctt 1102
 Db 425 ctgactattccggttaaacattacggaacaaagcctgacaacagcgaggttaaatggt 484
 Qy 1103 atttaactgataaataatggaatgtatcactogctcaggtacgaaatgactatattgtat 1162
 Db 485 taaaatctgtgacacactcaaaaatacaacgagtggaagatcacaaacgatttttgacg 544
 Qy 1163 ttgaagtgatggcttattactactcaaaacttatgataatggaagctactacaagaagt 1222
 Db 545 gagacgaaaaacatacatcagaacgttcagcgttatcgatgaagggaattatcatatccg 604
 Qy 1223 ccgataatattgcaatgcgtgatctcatgtaattgaagatggttaattgtgatcggtacc 1282
 Db 605 gcgacaaccatcgtgagagacctactacgtttgaaga---caaagggccataataacc 661
 Qy 1283 ttgttttgaagcaagtactggtttgaaaaattatcaaggcgagggaccataattataact 1342
 Db 662 ttgtattcgaagcaaacacaggggaacagaaacggatcaccaaggcggaagatctttatta 721
 Qy 1343 ggttaaatattgcgagatgacgcatttaataatcaagagcttattttagaattctttcca 1402
 Db 722 acaaacgctactacggtcggtcggtgacgaactcttccgttaagaagcagaagcttcagc 781
 Qy 1403 atgatgattataagatcggtgcaacttggtggttaagcagctatcggtatctctcaactaa 1462
 Db 782 agagcgttaaaaaacgcgtgctgagttagcgaacggtcgttcgttatcatagagtttaa 841
 Qy 1463 ataaggacgaaaagaatcctaaggtggtgagagttatgaattagtgataataataattactat 1582
 Db 893 cgtactgatgaatcgagcgcgcgaatgttttcaaaatgaacgcaaatggtactgtg 952
 Qy 1583 ttgcccgtaccgcgttttaaatcgaggaagtaagatgagctgttgatgaatgataatg 1642
 Db 953 tcaactgattcacgcgtttcaaaatgacgacgtggtatcaactcaacgatatatta-- 1010
 Qy 1643 cgttggtgataatgttgcgaatggtcggtatgttgcgtatgctgataacttggtacttata 1702
 Db 1011 -----catgctgttgatgatacaactcttcaacggttcaacggttaca 1051
 Qy 1703 agccattaaatgattctgtgagtagtctgacgtcttcttctgcaaacgtggtgacag 1762
 Db 1052 agccgtgacaaaacaggggtgtgctgcaaatgggtctgtatccaaacgagtgtgacat 1111
 Qy 1763 caacttattcatattatgctgtcccggttgcggaagaaagatgaccaagtattagttactt 1822
 Db 1112 tcaacttactctactctcagtcgagtcgcaagcc---aaaggcaacaatggtggttaca 1168
 Qy 1823 catatgactaataagaatggagtagcgggtgaagggaatgattcaacttgggcacga 1882
 Db 1169 gctacatgacaaacagaggtcttcttcgaggtataa-----aaggcaacatttggccaa 1222
 Qy 1883 gttttactacaaattaaacccggaataacacaaactactgttttagctaaatgactaatc 1942
 Db 1223 gcttttaataatgaacatcaaaaggcaataaaacatccgttgcataaaacagcatctcgtgagc 1282
 Qy 1943 aaggggattgattgg 1959
 Db 1283 aaggacagctgacatgg 1299

RESULT 10
 AAS85786
 ID AAS85786 standard; cDNA; 2079 BP.
 XX AAS85786;
 AC AAS85786;
 XX
 DT 13-FEB-2002 (first entry)

AAS88703/c
 ID AAS88703 standard; cDNA; 2079 BP.
 XX
 AC AAS88703;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #24507.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABC24516.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1; SEQ ID NO 24507; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2079 BP; 461 A; 475 C; 433 G; 710 T; 0 other;

	Query Match	3.0%	Score 70.2	DB 23	Length 2079
	Best Local Similarity	46.9%	Pred No. 2.7e-06		
	Matches 542	Conservative 0	Mismatches 553	Indels 60	Gaps 8
QY	817	gacaaagtgttaattggaatggtcatcaacttgcattgcgcgaatgatgggaattccaaac	876		
Db	1833	GGACAGTAGCTGAATCAACGGCTATCAGTTGTGTTGCTCTCGGGAACCGCGAAAC	1774		
QY	877	caaaatgataatcatatctatctctataataaagtatgggtgataatgaattaaagtcat	936		
Db	1773	AGCGTGTATGACATATCAATCTACATGTTTATCAAAAGCTGGAAAACGCGGGCGGTCTC	1714		

Qy	937	tggaagaatgtaggtccaatttttgctataattctacc-----gcggtttccaagaag	990
Db	1713	TTTAAAGACAGCGATAAAGTTCGACGCCAACGATCCGATCTCTGAAAGATACACAGCGCAAGAA	1654
Qy	991	tgttcaggatcaagctgtgttttgacagtgtaactctatcccaattattttatacaagggtta	1050
Db	1653	TGgtTCGGTTCtCAACCTTTACATCTGACGGAAATCCGGTTATTCTACAC-----T	1600
Qy	1051	gacagctctgataacaataccaatcatcaaaaaattgctagcgctactctttatttaact	1110
Db	1599	GACTATTCGGTAAACATTACCGGCAAAACAGCCCTGCAACACGCGCAGGTAAATGTGTCA	1540
Qy	1111	gataataatggaattgtatcacctcgctcaggtacagaaatgactatatgttatggaagt	1170
Db	1539	AAATCTGTATGACACACTCAAAATCAACGGATGGGAAGTCAAAAACGATTTTGTACGGA	1480
Qy	1171	gatggctattactaccaacttatgatcaatggga-----aagctactacaagaatgccc	1224
Db	1479	GACGGAANAACNATACGAACGTTTCAGCAGTTTATCGATGAAGCAATTTATACATCCGGC	1420
Qy	1225	gataattatgcaatcgctgtagtctcatgtaatgaaatggttaatggtatcggtaccctt	1284
Db	1419	GACAAACCATACGCTGAGAGACCTCACTACGTTGAAG-----CAAAGGCCATAAATACCTT	1363
Qy	1285	gttttgaagcaagtactggtttggaaaaattataagcgagagaccaattataactgg	1344
Db	1362	GTATTCGAAGCAACACGGGAACAGAAACGGATACCAAGCGCAAGAACTTTATTATAAC	1303
Qy	1345	ttaattatggcgagatgacgcgaatttaataatacaagagcttattagaattcttccaat	1404
Db	1302	AAAGCGTACTAGCGCGCGCGCAACGACTTCTTCGTAAAGAAAGCCAGAAAGTTCAGCAG	1243
Qy	1405	gatgatattaaagctggcgcaacttggtggttaattgcagctatcggtatctctcaactaat	1464
Db	1242	AGCGCTAAAAAACCGCATGTGTAGTTAGCGAACGGCGCCCTCGGTATCATAGAGTTAAAT	1183
Qy	1465	aagcgacaaagaatactcaaggtggcagagttatactaccattaatcttcgcaccaatg	1524
Db	1182	AATGATTACACATGGAANAAGTAATGAAG-----CCGCTGATCACTCAAAACACG	1132
Qy	1525	gtaagcgatgaattgagcgaccaaatgagttaaattaggtaataaataattactatttt	1584
Db	1131	GTAACTGATGAATCAGCGCGCGAATGTTTCAAAATGAACGGCAATGGTACTTGTGTC	1072
Qy	1585	gcgcgtaccgcgtttaaactcgagaagaatgatgatgtgttgatgaaatgctaattatgcc	1644
Db	1071	ACTGATTACAGCGGTTCAAAAATGACGATCGATGTGATTAATCTCAACGATATTTA----	1014
Qy	1645	gttggtgataatgttgcaatgctcggtatgttgcgtatgttgcgtatgcttaactggaat	1704
Db	1015	-----CATGCTTGTTATGTATCAAACTCTTAAACGGCGCCCTTCAACAG	973
Qy	1705	ccattaaatgattctggagtagcttgactgtctctctgcgcaaaactggcgacagca	1764
Db	972	CGCGTCACAACACAGGGCTTGTCGCAATGGGTCTTGATCCAAACGATGTGACATTC	913
Qy	1765	acttattcatattgctgtccccgttgcggaaaaagatgatccaactgttgccaccagt	1824
Db	912	ACTTACTCTCACTTCGCAGTCCCGCAGCC-----AAAGGCAACAAATGGGTATTCAACAGC	856
Qy	1825	tatagtcaataagaattggagtagcggttgaaaggaatggattcaactgttgccaccagt	1884
Db	855	TACATGACAAACAGAGGCTTCTTCGAGGATAAAA-----AGCAACATTTGGCGCAACG	802
Qy	1885	tctctacaaaattaccgccgataaacacactactgttttagctaaaaatgactaatca	1944
Db	801	TTCTTAATGAATCAACAGGCATATAAAGCATTAACATCCGTTGTCAAAACAGCATCTGGAGCA	742
Qy	1945	ggggattggattgg	1959
Db	741	GGACAGCTGACATGG	727

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 21, 2002, 23:37:00 ; Search time 12398.7 Seconds
(without alignments)
3995.025 Million cell updates/sec

Title: US-09-995-587A-2

Perfect score: 2367

Sequence: 1 atgtataaaagcggtaaaaa.....ttaacaaacgtcgatttaac 2367

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

1	2367	100.0	2367	6	AX316614
2	2367	100.0	2394	6	AX316615
3	2367	100.0	2592	6	AX316616
4	718.6	30.4	4305	1	STRETF
5	637	26.9	4634	6	AX316622
6	548	23.2	3600	1	STRETFB
7	97.8	4.1	11318	1	AE007686
8	84	3.5	155204	2	AC007926
9	80	3.4	2007	1	BSSACBG
10	80	3.4	5941	12	SYNPLRC
11	80	3.4	6347	12	SYNPLRL
12	80	3.4	6798	12	SYNPLRL
13	80	3.4	10317	6	AR162197
14	80	3.4	88063	1	BS294043
15	80	3.4	209510	1	BSUB0018
16	78.4	3.3	5831	12	AF047518
17	78.4	3.3	5842	12	AF004910
18	78.4	3.3	5846	12	XXU17500
19	78.4	3.3	6349	12	AF047519
20	78.4	3.3	9297	12	AF048702
21	78	3.3	180668	2	AC020857
22	76.4	3.2	10115	5	AF397467
23	75.2	3.2	172307	2	AC044842
24	75	3.2	22398	5	FRU271723
25	75	3.2	164399	5	PFMAL3P6
26	73.6	3.1	1422	1	BSU34874
27	73	3.1	35793	5	AY016024
28	73	3.1	154918	9	AL138781
29	72.6	3.1	666	8	AF413050S2
30	72.2	3.1	393	1	MBOVSP02
31	72.2	3.1	3669	1	AF396972
32	72.2	3.1	3693	1	AF396971
33	72.2	3.1	16136	1	AF396969
34	72.2	3.1	16204	1	AF396970
35	72	3.0	586	1	AF362688
36	71.4	3.0	144505	2	AL662884
37	70.8	3.0	199551	2	AC006281
38	70.8	3.0	237258	2	AL606925
39	70.4	3.0	2167	3	PRSTARPA
40	69.6	2.9	53870	3	AE003163
41	69.4	2.9	185994	2	AC002042
42	68.8	2.9	532	1	AF362685
43	68.8	2.9	550	1	AF362687
44	68.8	2.9	586	1	AF362690
45	68.8	2.9	3730	6	I34014

ALIGNMENTS

RESULT	1	AX316614	Sequence 2 from Patent WO0190319.	DNA	linear	PAT 14-DEC-2001
LOCUS	AX316614	Sequence 2 from Patent WO0190319.	2367 bp			
DEFINITION	AX316614	Sequence 2 from Patent WO0190319.				
ACCESSION	AX316614	Sequence 2 from Patent WO0190319.				
VERSION	AX316614.1	GI:17899728				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
Source						
BASE COUNT						
ORIGIN						

QY 2101 acgccaactacacagaaacccctgagacacctaataactccccaaacacccaaagactcct 2160
Db 2101 ACGCCAATACACCAAGAACCCCTGAGACACCTAATACTCCCAAAACACCAAGACTCCT 2160
QY 2161 gaaaatcctgggacacccctcaaaactcctaatacacctaataactccggaattcctttaact 2220
Db 2161 GAAAATCCTGGGACACCTCAAACTCCTAATACACCTAATACTCCGGAATTCCTTTAACT 2220
QY 2221 ccagaaacgcctaagcaacaccccaaaactaaataatcgcttggcccaaaactggaat 2280
Db 2221 CCAGAAACGCCTAAGCAACCTGAAACCCCAAACTAATAATCGTTGCCACAAACTGGAAT 2280
QY 2281 aatgccaaataagccatgatggcctagcttagtggaacactgcttagtgattggtctt 2340
Db 2281 AATGCCAATAAGCCATGATGGCCTAGCTAGTATGGGAACATGCTTAGTATGTTGGTCTT 2340
QY 2341 gcagaaattaaacacgcctgattaac 2367
Db 2341 GCAGAAATTAAACACGTCGATTTAAC 2367
RESULT 2
AX316615
LOCUS AX316615 2394 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 3 from Patent WO0190319.
ACCESSION AX316615
VERSION AX316615.1 GI:17899729
KEYWORDS
SOURCE Lactobacillus reuteri.
ORGANISM Lactobacillus reuteri.
REFERENCE Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
AUTHORS I (sites)
Hijum, S.A.
TITLE Novel fructosyltransferases
JOURNAL Patent: WO 0190319-A 3 29-NOV-2001;
TWO (NL) Location/Qualifiers
source
1. .2394
/organism="Lactobacillus reuteri"
/db_xref="taxon:1598"
BASE COUNT 829 a 437 c 464 g 664 t
ORIGIN
Query Match 100.0%; Score 2367; DB 6; Length 2394;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtataaaacggtataaaattggcgagtcgtttacactctcgactgctgcgtgattt 60
Db 28 ATGTATAAAACGGGTAAAAATTGGCGAGTCGTTACACTCTCGACTCTCGCTGCTGCTGATTT 87
QY 61 ggtgcaacaactgtataatgcatccgcgacacaaataattgaaacaatgattcttact 120
Db 88 GGTGCACAACTGTAAATGCAATCCCGGACACAAATATTGAAAACAATGATTCTTCTACT 147
QY 121 gtacaagttacaacaggtgataatgatattgctgttaaaagtgtgacactggtagtggt 180
Db 148 GTACAAGTTACAACAGGTGATAATGATATTGCTCTTAAAAGTGTGACACTTGGTAGTGGT 207
QY 181 caagttagtgagtagtagtacactattagaactctgctaataatgcaatagtcttct 240
Db 208 CAAGTTAGTCAGCTAGTGTATACACTATTAGAACTTCTGCTAATGCAAAATAGTCTCT 267
QY 241 tctgcccttaatacacaaaattctaacagtcagtcagtagcaagttctgctgaataacatca 300
Db 268 TCTGCCGCTAATACAAAAATTCCTAACAGTCAAGTAGCAAGTCTGCTGCAATTAACATCA 327
QY 301 tctacaagttccgcagcttctattaaataacacagatagtaagcgctcaagaaaataact 360
Db 328 TCTACAAGTTCGCGAGCTTCTATTAAATTAACACAGATAGTAAGCGGCTCAAGAAATACT 387

QY 361 aatacagccaaaatgatgacagcaaaaaagctgcaccagctaacgaatctcttgaagct 420
Db 388 AATACAGCCAAAATGATGACACGCAAAAAGCTGCACCAGCTAACGAATCTTCTGAAGCT 447
QY 421 aaaaaagacccagctgtataacgttaattcttctcagctgcaaaaaatgatgatcaacaa 480
Db 448 AAAAATGAACAGCTGTAAACGTTAATGATCTTTCAGCTGCAAAAATGATGATCAACAA 507
QY 481 tccagtaaaaagaataactaccgctaagttaacaagagatgctgaaaacgttgttaaaaaag 540
Db 508 TCCAGTAAAAAAGAAATACTACCGCTAAGTTAAACAAGGATGCTGAAAAACGTTGTAAAAAAG 567
QY 541 gcgggaattgatcctaacagctttaactgatgaccagattaaagcattaaataaagatgaac 600
Db 568 GCGGGAATTGATCCTTAACAGTTTAACTGATGACCAAGATTAAAGCATTAATAGATGAAC 627
QY 601 tctcgaagctgcaaaagctgtgttacacaaaatgacttataatgattcccaaaagattgct 660
Db 628 TTCTCGAAAAGCTGCAAAAGTCTGTTACACAAATGACTTATAATGATTTTCCAAAAGATTGCT 687
QY 661 gatacgttaatacaacaagaatgctgacacagttccattctttaaagcaagtgaatc 720
Db 688 GATACGTTAATCAACAAGATGTCGGTACACAGTTCCATTTCTTTAAAGCAAGTGAATC 747
QY 721 aaaaatagctcctgcctcaactaaagatgcacaaactaatactattgaaccttttagat 780
Db 748 AAAAATATGCTGCGCTACAACTAAAGATGCACAACTAATACTATTGAACCTTTAGAT 807
QY 781 gtatggatcatcggcagttcaagatgttcggacaggaacaagtgtcgaattggaatggc 840
Db 808 GTATGGGATTCATGGCGCAGTTCAAGATGTCGGACAGGACAAAGTGTGTAATTTGGAATGGC 867
QY 841 tatcaactgtcatcgcaatgatggaaattccaaacaaatgataatacatatctc 900
Db 868 TATCAACTTGTCTATCGCAATGATGGGAATTTCCAAACCAAAATGATAATCATATCTATCTC 927
QY 901 ttataataaagtatggtgataaataagtcatttggaagaatgtaggccaattttt 960
Db 928 TTATATAATAAGTATGTTGATATAATGAATTAAGTCATTTGGAAGAAATGTAGTCCAAATTTT 987
QY 961 ggcataatctaccgcggtttcacagaatggtcaggatcagctggttttgaacagtgat 1020
Db 988 GGCTATAATTCTACCGCGGTTTCCACAGAATGTCAGGATCAGCTGTTTGAACAGATGAT 1047
QY 1021 aaactctaccaattatttatcaagggtagacacgtctgatacaacaatacaaatcatcaa 1080
Db 1048 AACTCTATCCAAATATTATACAAAGGTAGACACGCTCTGATAACAATAACCAATCATCAA 1107
QY 1081 aaaaatgctagcgtactctttatttaactgataataaaggaaatgtatcacctcgctcag 1140
Db 1108 AAAATTTGCTAGCGCTACTCTTTATTAACTGATAAATAAGGAATGTATCCTCAGCTCAG 1167
QY 1141 gtacgaatgactatattgtattggaagtgatggtgttactaccacaaacttatgatcaa 1200
Db 1168 GTACGAATGACTATATTGTTGAAGGTGATGGCTATTACTACCAAACTTATGATCAA 1227
QY 1201 tggaaagctactcaacaagaagtgccgataattgcaatgctgtagctcatgaattgaa 1260
Db 1228 TCGAAAGCTACTTAACAAGGTGCCGATAATATTGCAATGCGTGTGCTCATGTAATTTGAA 1287
QY 1261 gatggttaattggtgatcggtacacctgtttttgaagcaagtaactggtttggaaaattatcaa 1320
Db 1288 GATGTTAATGTTGATCGGTACCTTGTTTTGAAGCAAGTACTCGTTTGGAAATATATCAA 1347
QY 1321 ggcaggagcaaaatttataaactggtttaactatggtcggagatgacgacattataaag 1380
Db 1348 GCGGAGGACCAAAATTTATAACTGGTTAAATTTATGGCGGAGATGACGCAITTTAATATCAAG 1407
QY 1381 agcttatttagaatcttctccaatgatattaaagtcgggcaactgggctaataatgca 1440
Db 1408 AGCTTATTAGAAATCTTTCCAAATGATGATATTAAGAGTGGGCAACTTGGGCTTAATGCA 1467

```
QY 1441 gctatcggtatctctcaaaactaaataaaggagcgaagaaatccttaagtggtgagagttatatac 1500
Db 1468 GCTATCGGTATCTCTCAAACTAAATAAGGAGCAGAAAGAACTCTAAGTGGCGAGAGTTATAC 1527
QY 1501 tcacattattctgcacacaaatggaagcgtatgaatgagcgacaaatgtagttaaa 1560
Db 1528 TCACCAATTAATTTCTGCACCAATGTTAAGCGATGAATTTGAGCGACCAATGTTAGTTAAA 1587
QY 1561 tttagttaaataattactatttcccgctacccgttttaaatcgcaggaagtaagtatgat 1620
Db 1588 TTAGGTAAATAATATTTACTTTTGGCGCTACCCGTTTAAATCGAGGAAGTAATGATGAT 1647
QY 1621 gcttgatgaatgctaattatccgttggtgataatggttgcaatggtcgagatggtgct 1680
Db 1648 GCTTGGATCAATGCTAATTTATGCCGTGTGTGATGATGTTGCAATGTCGCGATGTTGCT 1707
QY 1681 gatagtcaactggatcttataagccattaaatgattctggagtagcttgcagctgtct 1740
Db 1708 GATAGTCTAACTGGGATCTTTAAGCCATTAAATGATTCTGGAGTAGTCTTGAGCTGCTCT 1767
QY 1741 gttcctgcaaaactggcgagacagcaacttattcatattatgctgtcccggttccggaaaa 1800
Db 1768 GTTCTGCAAACTGGCGACACCACTATTATCATATTATGCTGTCTCCCGTTCGCGGAAAA 1827
QY 1801 gatgaccaagtatttagttacttcataatgactaataatgaaatggaatgagtagcggttaaagga 1860
Db 1828 GATGACCAAGTATTAGTTACTTCATATGATGACTAATAGAAATGGAGTAGCGGGTAAGGA 1887
QY 1861 atggattcaacttggccacggagttcttactacaataacccgggatacaacaactact 1920
Db 1888 ATGGATTAACACTTGGGACCCGAGTTCTTACTACAAATTAACCCGGATTAACCAACTACT 1947
QY 1921 gttttagctaaaatgactaaatcaaggatggtgattggagtagtcaagcgaataatctt 1980
Db 1948 GTTTTAGCTAAATGACTAATCAAGGGATGGATTGGGATGATTCAAGCGGAAATCTT 2007
QY 1981 gatagtattggtgattagactccgctgtcttaccctggcgaaacgtgataaaactgttgat 2040
Db 2008 GATATGATTGGTGTGATTAGACTCCGCTGCTTACCTGGCGAACGTGATAAACCTGTGTAT 2067
QY 2041 tgggacttaattggtatggtatgaataacccgatgactcgtctacacccaaatgactcgtgaa 2100
Db 2068 TGGGACTTAATGTTATGGATTAAACCCGATGATCTCTGCTACACCAAAATGATCCTGAA 2127
QY 2101 acgcaactacacagaaacccctgagacacacttaatactcccaaaacaccaaagactcct 2160
Db 2128 ACGCAACTACACCAAGAAACCCCTGAGACACCTAATACTCCCAAAACACCAAGACTCCT 2187
QY 2161 gaaaactcggagacacactcctaaactcctaaacacacttaatactcctcggaaattccttaact 2220
Db 2188 GAAATCTGGGACACCTCAAACTCCTAATAACACTAATACTCCTCGGAAATTCCTTTAACT 2247
QY 2221 ccagaaacgcttaagcaacctgaaccccaactaataatcgttttgccacaaactggaat 2280
Db 2248 CCAGAAACGCTTAGCAACCTGAACCCAACTAATACTGTTGCCAACAACTGGAAT 2307
QY 2281 aatgccataaagccatgattggcctaggatgggaacattgcttagtattgttgcttt 2340
Db 2308 AATGCCAATAAAGGCATGATTGGCCTAGGTATGGGAACATTGCTTAGTATGTTTGGTCTT 2367
QY 2341 gcagaaattacaacgctcgatttaac 2367
Db 2368 GCAGAAATTAACAACGTCGATTTAAC 2394

RESULT 3
LOCUS AX316616 2592 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 4 from Patent WO0190319.
ACCESSION AX316616
VERSION AX316616.1 GI:17899730
KEYWORDS Lactobacillus reuteri.
```

```
ORGANISM Lactobacillus reuteri
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
Lactobacillus.
REFERENCE 1 (sites)
AUTHORS van Geel-Schutten,G.H., Rahaoui,H., Dijkhuizen,L. and van
Hijum,S.A.
TITLE Novel fructosyltransferases
JOURNAL Patent: WO 0190319-A 4 29-NOV-2001;
TNO (NL)
FEATURES
source Location/Qualifiers
1..2592
/organism="Lactobacillus reuteri"
/db_xref="taxon:1598"
misc_signal 1..67
/Note="Putative expression-regulating region"
RBS 29..32
RBS 54..57
misc_signal 2438..2592
/Note="Putative expression-regulating region"
BASE COUNT 896 a 465 c 501 g 730 t
ORIGIN
Query Match 100.0%; Score 2367; DB 6; Length 2592;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2367; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 1 atgtataaaagcgtgtaaaattggcagtcgtttacactctcgactctcgactgctcgctggtattt 60
Db 68 ATGTATAAAAGCGGTAAATTTGGGCAGTCTGTACTCTCGACTCTCGCTGCGCTGGTATTT 127
QY 61 ggtgcacaactgtaaatgcatccgcggacacaaatattgaaaaaatgattcttctact 120
Db 128 GGTGCACAACCTGTAATGCTATCCGGGACACAAATATTGAAAACAATGATTCTTCTACT 187
QY 121 gtacaagttacaacaggtgataatgattgtgtttaaagtgtgacacttgtagtggt 180
Db 188 GTACAAGTTACACACAGTGATATGATTGTCTGTAAAGTGTGACACTTGGTAGTGGT 247
QY 181 caagttagtcagctagtgatcacgactatttagaactctctctaagcaaatagtcctct 240
Db 248 CAAGTTAGTCAGCTAGTGATGACTATTAGAACCTCTCTAATCAATAGTGCTCTCT 307
QY 241 tctgcgcctaaatacacaattcttaacagtcgaagtagcaagtcgtcgtcgaatacatca 300
Db 308 TCTGCGCTAATACACAAATTTCTAAAGTCAAGTAGCAAGTTCTGCTGCAATAACATCA 367
QY 301 tctacaagttccgcagcttcaattaaatacacagatagtaaaagcgttaacgaatact 360
Db 368 TCTACAAGTTCCGCAGCTTCATTAAATAACACAGATAGTAAGCGGCTCAAGAAAAATACT 427
QY 361 aatacagccaaatgatgacacgcaaaaagctgcaccagctaacgaattcttctgaagct 420
Db 428 AATACAGCCAAAATGATGACACGCAAAAAGCTGACCAAGCTAACGAACTCTCTGAAGCT 487
QY 421 aaaaatgaaccagctgtaaaccttaattgtcttcagctgcaaaaaatgatgatacaaa 480
Db 488 AAAAATGACCAAGCTGTAAACGTTTAATGATCTTTCAGCTGCAAAAATGATGATCAACAA 547
QY 481 tccagtaaaagaatactactaccgcgttaagttaaacaagatgctgaaacgcttgtaaaaaag 540
Db 548 TCCAGTAAAAACAATACTACCGCTTAAGTTAAACAAGGATGCTGAAAACGCTGTGAAAAAG 607
QY 541 gcgggaattgatccttaacagtttaactgatgaccagattaaagcattaaataagatgaac 600
Db 608 GCGGGAATTGATCTCAACAGTTTAACATGATGACCAAGTTAAAGCATTAATAAGATGAAC 667
QY 601 tcttcgaagctgcaagctctggttacacaaatgactataatgatttcacaaagattgct 660
Db 668 TTCTCGAAGCTGCAAGTCTGCTACACAAATGACTTATATGATTATCCAAAAGATTGCT 727
QY 661 gatacgttaatacaacaagatggtcgggtacacagttcccatctttaaagcaagtgaaatc 720
|||||
```

```
Db 728 GATACGTTAATCAACAAGATGGTCGGTACACAGTTCCATTCTTTAAAGCAAGTGAAATC 787
Qy 721 aaaaatgctgcgcgtacaaactaaagatgcacaaactaatactattgaacctttagat 780
Db 788 AANAATATGCTGCGGTACAACTAAAGATGACAAACTAATACTATTGAACCTTTAGAT 847
Qy 781 gtatggattcatgcccagttcaagatgctcgacagacaaagtgttaattggaatggc 840
Db 848 GTATGGGATTCATGCCAGTTCAAGATGTTGCGGACAGACAAGTTGCTAAATGGAATGGC 907
Qy 841 tatcaactgtcatcgcaatgatgggaattccaaaccaaataatgataatcatatctctc 900
Db 908 TATCAACTTGTTCATCGCAATGATGGGAATCCAAACCAAAATGATATCATATCTATCTC 967
Qy 901 ttataataaagatgataatgaataatgaatcattgggaagatgtagtccaattttt 960
Db 968 TTATATAATTAAGTATGATGATTAATTAAGTCAATTTGGAAGAAATGATGCTCCAATTTT 1027
Qy 961 ggcataattctaccgcyggtttcacagaatggtcagatcagctgttttgaacagtgat 1020
Db 1028 GGCATATAATTTCTACCGGGTTTCACAAGATGTCAGGATCAGCTGTTTGAACAGTGAT 1087
Qy 1021 aactctatcaattattttatatacagggtagacacgctctgtatacaataaccaatcatcaa 1080
Db 1088 AACTCTATCCAAATATTTTATACAAGGGTAGACACGCTCTGATTAACAATACCAATCATCAA 1147
Qy 1081 aaattgtagcgtactctctttttaaactgataataatggaatgtatacaactcgctcag 1140
Db 1148 AAAATTGCTAGCGCTACTCTTTATTTAACTGATTAATGGAATGATGATGCTGCTGCTCAG 1207
Qy 1141 gtacgaatgactatttbtattggaagtgtggtctattactaccaaaacttatgatcaa 1200
Db 1208 GTACGAATGACTATATTTGTTGAAGGTGATGCTGCTATTACTACCAACTTATGATCAA 1267
Qy 1201 tggaaagctactaaacaaagtcgcaataattgcaatgcgtgagctcatgtaattgaa 1260
Db 1268 TGGAAAGCTACTAACAAAGTGCCGATAATATGCAATGATGCAATGCTGCTGCTGCTGCTG 1327
Qy 1261 gatgtaattggtgagtggttacctgtttttggaagcaagtaactggttggaaattatcaa 1320
Db 1328 GATGTAATGTTGATCGGTACCTGTTTGAAGCAAGTACTGTTTGGAAATTAATCAA 1387
Qy 1321 ggcgagaccacaaattataactggttaattatgacgagatgacgcatttaataatcaag 1380
Db 1388 GCGGAGGACCAATTTATAACTGGTTAAATTTATGCGGAGATGACGCATTTAATATCAAG 1447
Qy 1381 agcttatttagaattcttcccaatgatataaagtcgagacgcaactggcctaagca 1440
Db 1448 AGCTTATTAGAATCTTTCCAAATCATGATATAGAGTCGGGCACTTGGGCTAATGCA 1507
Qy 1441 gctatcggtatctcctcaactaaataagacgcaaaagaaatccttaagtgagagttatc 1500
Db 1508 GCTATCGGTATCCTCAAACTAATAAAGACGAAAGAAATCCTAAGGTGGCAGATTATAC 1567
Qy 1501 tcaccattaattctgcaccaatgtaagcatgaaatgagcagacaaatgtagttaaa 1560
Db 1568 TCACCAATAATTTCTGCACCAATGTAAGCATGAAATTTGACGACCAAAATGTAGTTAA 1627
Qy 1561 ttaggtaataatattactatttgcgcctaccggtttcaaatcagaggaagtaagatgat 1620
Db 1628 TTAGGTAATAAATATTACTTTATTTGCCGCTACCCGTTTAAATCGAGGAAGTAATGATGAT 1687
Qy 1621 gcttggaatgataatattgacctgctaccggtttcaaatggtcggaatgattgct 1680
Db 1688 GCTTGGAATGAATGCTAATTTATGCGGCTTGGTGATAATGTTGCAATGGTGGGATATGTTGCT 1747
Qy 1681 gatagctcaactgagctctataaagcattaaatgattctgagagtagcttgactgctct 1740
Db 1748 GATAGCTAAGTCTTATTAAGCCATTAATGATTTCTTGAGTAGTCTTGACTGCTTCT 1807
Qy 1741 gttcctgcaactggcgagacgaacttattcatattatgctgtccccgttgcgggaaaa 1800
Db 1808 GTTCTGCAAACTGGCGGACAGCAACTATTATCATATTATGCTGTCCCCGTTGCCGGGAAAA 1867
```

```
Qy 1801 gatgaccaagtatttagttacttcatatgactaataagaaatggagtcgagggtaaaaga 1860
Db 1868 GATGACCAGATTAATAGTTACTTATATATGACTAATAGAAATGGAGTAGCGGGTAAAGGA 1927
Qy 1861 atggattcaacttggccaggtttcttactacaaaattaaacccggataaacacactact 1920
Db 1928 ATGGATTCAACTTGGGACCGGATTTCTTACTACAAATTAACCCGGATAACACAATACT 1987
Qy 1921 gttttagctaaatgactaatacaagggttggatttgggtgatttcgaagcgaaatctt 1980
Db 1988 GTTTTAGCTAAATGACTAATCAAGGGATTTGGATTGGGATGATTCGAAGCGAAATCTT 2047
Qy 1981 gatgattggtgatttagactccgctgtctttacctggcggaactgataaacctgttgat 2040
Db 2048 CATATGATTGTTGATTTAGACTCCGCTGCTTTTACCTGGCGAACGTTGATAAACTCTT 2107
Qy 2041 tgggacttaattggttattgattaaacccgcatgctctgctacacaaatgactcgtgaa 2100
Db 2108 TGGGACTTAATTTGTTATGGATTAAACCGCATGATCTCTGCTACCAAAATGATCCTGAA 2167
Qy 2101 acgccaactacacagaaaccccttgagacactaactcccaaaacacacaaagactcct 2160
Db 2168 ACGCAACTACACAGAAACCCCTTGAGACACTAATACTCCCAAAACACCAAGACTCCT 2227
Qy 2161 gaaaactcctgggacacctcaaacctccttaataacacactaactcctcggaattcctttaa 2220
Db 2228 GAAATCCTGGGACACCTCAAACTCCTAATACACTAATCTCGGAAATTTCTTTAACT 2287
Qy 2221 ccggaacgctgaacacactgaaaccccaactaataatcgtttgcacaaactggaaat 2280
Db 2288 CCAGAAACGGCTTAGCAACCTGAAACCCAACTAATAATGTTTGGCCAAACTTGGAAAT 2347
Qy 2281 aatgcgaataaagccatgattgacctaggtatgggaacattgcttagtattggtctt 2340
Db 2348 AATGCCAATAAAGCCATGATTGGCTAGGTATGGGAACATTTGCTTAGTATGTTGGTCTT 2407
Qy 2341 gcggaattaaacaaacgtcgatttaac 2367
Db 2408 GCAGAAATTAACAAACGTCGATTAAAC 2434
```

RESULT 4

```
STRTF S.mutans fructosyltransferase 4305 bp DNA linear BCT 15-FEB-1996
DEFINITION M18954
ACCESSION M18954.1 GI:153635
VERSION fructosyltransferase.
KEYWORDS Streptococcus mutans (strain GS5) (clone: pSS22 and pTS102.) DNA.
SOURCE Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ORGANISM Streptococcus.
REFERENCE 1 (bases 1 to 4305)
AUTHORS Shiroza, T. and Kuramitsu, H.K.
TITLE Sequence analysis of the Streptococcus mutans fructosyltransferase
JOURNAL gene and flanking regions
MEDLINE J. Bacteriol. 170 (2), 810-816 (1988)
COMMENT Draft entry and printed copy of sequence for [1] kindly provided by
H.K. Kuramitsu, 06-MAR-1988.
FEATURES
Source Location/Qualifiers
1..4305
/organism="Streptococcus mutans"
/strain="GS5"
/db_xref="taxon:1309"
/clone="pSS22 and pTS102."
26..394
/note="ORF1; putative"
/codon_start=1
/transl_table=11
/product="unknown protein"
/protein_id="AA88583.1"
CDS
```


QY	948	agg	tccaatttttggcgtataaattctctaccgcgggtttccaaagaattgctcagatcagctgt	1000
Db	2107	AGGT	TCAATCTTTTGGAACTAAAGAAACAAATGTTTCCAAGATGGTTCAGGTTCAAGTAT	2166
QY	1008	ttg	aacagtgataaactctctatccaattatcttatacacaaggtgacacgctctgtataacaa	1067
Db	2167	TGTA	AAATGATGATGGTACAAATCAACTATTTTTCACCTCAAAATGATACGTCTGATTAACA	2226
QY	1068	tacca	atcatcaaaaaattctctagcgtactcttttatttaactgataataatggaatgt	1127
Db	2227	GTTGA	ATGATCAACGCCTTGCACCGCAACATTAACCTTAATGTTGATGATAACGGTGT	2286
QY	1128	atac	ctcgctcaagtagcagaatagctatatattgtatttgaaggtgatggcttactacca	1187
Db	2287	TTCA	TCAGAGTGTGTAAATATATCAAGTTTGTGTTTGAAGGTGATGGATTCAC	2346
QY	1188	aact	tatgatcgaatggaaagcTactCaacaaggtg-----ccgataaatattgcaatgggtga	1244
Db	2347	AAC	TATGAACAATTTCGCAACCGCAAGATCGTGAAATGATGATTACTGCTTACGTGA	2406
QY	1245	tgct	aatgtaattgaagtagtgaattggtgctgcgttaocttgttttgaagcaagtagctgg	1304
Db	2407	CCCA	CACGTTGTCAATATAGAAAAATGGTGATCGTATCTTGTATTCCGAAGCTAATACTGG	2466
QY	1305	tttgg	aaaaattatcaagcgcagaccaaattataaactgggttaaatattggcggagatga	1364
Db	2467	GACAG	AAGATTAACCAAAAGTCGACCAAAATTAATTTGGGCTAACPTATGGTGGCGATGA	2526
QY	1365	cgatt	aatatacaagagcttattagaattctttccaatgatgatattaaagtagtcggc	1424
Db	2527	TGC	TTTCAATATTAAGAGTTCCTTCAAGCTTTTCAATAATAAGAGGATCGTGAATTGGC	2586
QY	1425	aact	tggtcaatgcagctatcggtagctctcaaaactaaaaaagaagcaagaatccttaa	1484
Db	2587	TGG	TTTAGCTAATGTCGACTTGGTATCTTAAAGCTCACTAACAAATCAAGTAAGCAAA	2646
QY	1485	ggtg	cagagttataactcaccattaaattctgcaccaatggtgaagcgatgaaatggacg	1544
Db	2647	GGT	TGAAGAAGTATACTCACCATTGGTATCTACTTTGATGGCTTGCAGTACAGGTA-----	2701
QY	1545	accaa	atgattagtaataaataattacttatttgcgcgtaccctgttaaactcg	1604
Db	2702	-----	NNNNNAAGCTTGGTGATAAGTATATCTTCTCCGTAACTCGTGAAGTCG	2754
QY	1605	agga	agtaatgatgactgcttgatgaaatgctaaattatgcgcttggtgataaattgttgc	1664
Db	2755	TGG	TTCGATCGTGAATTAACCGCTAAGGATAACACAACTGTTGGTGATTAACGTTGCTAT	2814
QY	1665	ggtc	ggatatattgctgatagctaaactggatcttataaagccoataaattgattctggagt	1724
Db	2815	GA	TGTTACGTTTCCGATAGCTTAATTTGGGTAAAGTACAAAGCCATTAAATAACTCAGGTGT	2874
QY	1725	agct	ttagctctctctgcttcctgcgaactggcgacagcaacttattcatattatgctgt	1784
Db	2875	CGT	ATTAACGACTCAGTACTCGAACTGGCGTACTGCTACTTATCTCTACTATTCGACGT	2934
QY	1785	cccc	ttccggaataaatgaccaagtattagttacttcatatatactagtaataatgaaatgg	1844
Db	2935	ACCT	TAGCTGGTCTATCTGATCAAGTATTAATTAATTTACTTCTTACATGAGTAAACAAGACTT	2994
QY	1845	agta	cggtgaagaagaatggatccaacttggcgacagattcttcttactacaaattaaacc	1904
Db	2995	TGC	TTACGTTGAAGGAACATATGCACTTGGGCACCAAGTTTCTTAGTACAAATCAATCC	3054
QY	1905	ggata	acacaactactgttttagctaaaaatgactaatcaaggggattggatttgggtga	1964
Db	3055	AGAT	CACACGACAACACTGATTAGCAGTGCACACTAACCAAGGTGACTGGGTCTGGGACCA	3114
QY	1965	ttca	agggaaaaattcgaatgattggtagtt-----agactccgctgc	2009
Db	3115	CTCT	AGTCGGAAGCATATATGCTCGGTCTTCTTAAGAAGAGGTGCAGTAACAGTGC	3174

QY	2010	tttaactgggaacgtgataaacctgtgtgattgggactttaat-----tggtta	2057
Db	3175	CTTACCAGGTGAATGGGGTAAAGCCAGTTTGACTGGAGTTTGATTAAACAGAAAGTCTCGCTT	3234
QY	2058	tgattataaacgcgatgctgtctacaccaaataatctctgaaacgcgaactacaccaga	2117
Db	3235	AGGCTTAAAGCCTCATCAACCAAGTTCAACCAAAAGATTGAT-----CAACCTGATCAACA	3288
QY	2118	aacccttgagacacctaatactcccaaacacccaagaactcctcctgaaatcctctgggacacc	2177
Db	3289	ACCTTCTGGTCAAAACACTAAGAAATGTCACACCAGTGAATGCTGATAGCCTGCTGGTAA	3348
QY	2178	tcaaatcttaatacacataatactactcggaaataattccttaactcaag	2224
Db	3349	GGCAACTCTGTATACACTAATATTGATCCAAAGTCACAAAGCTTCTG	3395
RESULT	6		
STRETFB		STRETFB	
LOCUS		3600 bp	linear
DEFINITION		Streptococcus salivarius fructosyltransferase (ftf) gene, complete cds; uracil phosphoribosyltransferase gene, 5' end; IS1161	BCT 12-FEB-2001
ACCESSION		L08445	
VERSION			
KEYWORDS			
SOURCE		L08445.1 GI:348000	
ORGANISM		Streptococcus salivarius.	
REFERENCE		Streptococcus salivarius.	
AUTHORS		Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.	
TITLE		1 (bases 1 to 3600)	
JOURNAL		Rathsam,C., Giffard,P.M. and Jacques,N.A.	
MEDLINE		The cell-bound fructosyltransferase of Streptococcus salivarius: the carboxyl terminus specifies attachment in a Streptococcus gordonii model system	
REFERENCE		3 Bacteriol. 175 (14), 4520-4527 (1993)	
AUTHORS		2 (bases 1 to 3600)	
TITLE		Jacques,N.A.	
JOURNAL		Direct Submission	
FEATURES		Submitted (04-JAN-1993) Institute of Dental Research, Australia	
Source		Location/Qualifiers	
		1..3600	
		/organism="Streptococcus salivarius"	
		/db_xref="taxon:1304"	
		/tissue_lib="lambda-A39/ATCC 25975"	
		<1..56	
		/standard_name="IS1161"	
		/function="transposition"	
		/note="ORF4; putative"	
		/codon_start=3	
		/transl_table=11	
		/product="transposase"	
		/protein_id="AAA71924.1"	
		/db_xref="GI:348001"	
		/translation="IHGYQSAKLFLQTQA"	
		220..3129	
gene		/gene="ftf"	
		220..3129	
		/gene="ftf"	
		/function="sucrose cleavage and fructan synthesis"	
		/codon_start=1	
		/transl_table=11	
		/evidence=experimental	
		/product="fructosyltransferase"	
		/protein_id="AAA71925.1"	
		/db_xref="GI:348002"	
		/translation="MDITVNSQNSTVAPKQAECKMRYISIRKVTATGATSLVGLTAP	
		LCATGVKADQVETAPATATATPTSTASLTATSETATSVATSESSVVAHSEVA	
		TKPTVETQSTPTSPVSEKASSTVTVSSSDATTSAVAAVSAPHTSEAAVEAPT	
		TASSPRAATHFEDVLKVSNGAANLKNLGRKISVIEENMTSDQIVALTETEEIKAL	
		NKVFSDDAIGTSLTLYTRNLKLDIVAFSLKQDSKLVPYFKADTIINPAFTVDAQ	
		TMKDESDWSPQVQDAKSGVYSNNWQYLVISNAGAPNKNNSHIYLLRYKYGDND	

Db 2341 GACTTTAAATGATATTATATCGGTATCGGGTTATGCTTAAACACATACCTATCCT 2400
 Oy 2083 acacaaatgatcc 2096
 Db 2401 ACAGTTGATGTTTC 2414

RESULT 7
 AE007686
 LOCUS
 DEFINITION
 11318 bp DNA linear BCT 27-JUL-2001
 Clostridium acetobutylicum ATCC824 section 174 of 356 of the
 complete genome.
 ACCESSION
 AE007686 AE001437
 VERSION
 AE007686.1 GI:15024739
 KEYWORDS
 Clostridium acetobutylicum.
 ORGANISM
 Clostridium acetobutylicum.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 Clostridium.

REFERENCE
 AUTHORS
 1 (bases 1 to 11318)
 Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q.,
 Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y. I.,
 Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P.,
 Daly, M. J., Bennett, G. N., Koonin, E. V. and Smith, D. R.
 Genome Sequence and Comparative Analysis of the Solvent-Producing
 Bacterium Clostridium acetobutylicum
 J. Bacteriol. 183 (16), 4823-4838 (2001)
 21359325
 11466286

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

FEATURES
 source
 1. 11318
 /organism="Clostridium acetobutylicum"
 /strain="ATCC 824"
 /db_xref="ATCC:824"
 /db_xref="taxon:1488"
 89..790
 /gene="CAC1770"
 89..790
 /gene="CAC1770"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase sigma factor, SigK-like"
 /protein_id="AAK79735.1"
 /db_xref="GI:15024740"
 /translation="MLNRSLEDRIQAKESKEELNRLIOEYKPFIAISLQKKTGYL
 QYGDDELSTGIAFEAVESYDRNKGKFLSFAKRIILRAIDYIRKQKFRNIVLQ
 GNYDENNDINPINFNKAQYKNSINEIRLEILEYKELREWGTEFVNLVQASP
 QELKKTKNIANIVYHNQDILNKLIRVKRLPIKEIKNIYINRKLRLGRYIIAY
 IAEIGDYRLIREVIE"
 801..1544
 /gene="CAC1771"
 801..1544
 /gene="CAC1771"
 /codon_start=1
 /transl_table=11
 /product="Uncharacterized protein, ykri B. subtilis
 homolog"
 /protein_id="AAK79736.1"
 /db_xref="GI:15024741"
 /translation="MRGIVIEELKYKLIILTEKGFIDKLNTRAKIGOEINIKSKK
 ISTKALRRPALAATAILFFIINLIYGMITREYVVDINPNNKNVAMETHYVIA
 GNTIKLQAANKGVSTVKKRNLKFKATVWNNFRTAQNDKFDISGKQNTIVITIA
 SSKSINDESIOSSLEHYIKENKINARPMVLGNDIDYKKSQVGIPIDKFILINKVI
 KNPYSKYFDLNNKSIDELINANQENY"
 1682..3151
 /gene="CAC1772"
 /note="sacB1"

CDS
 gene
 CDS
 gene
 CDS

1682..3151
 /gene="CAC1772"
 /codon_start=1
 /transl_table=11
 /product="Levansucrase"
 /protein_id="AAK79737.1"
 /db_xref="GI:15024742"
 /translation="MNKLKIVKICILIGSMICSGIITQOTFASTDMNTKETYGVSHIT
 RYNSKIPMEONDLKFKYPOFNASTLKNIASAGYDKNGKLNLDLDVDSWPLQNGDGT
 VANYGCHLVFALAGPKNQDDTSIYMFYQKIGENSIDSWKAGVFKDSDKRVANDP
 YLKYQTQEWSSAGATILSDGQVRLFTYDFSGVAKDGGTDASNOVITTVNLSQPSNT
 INIDSVDHKSVPDGGNETIYQNVQOFIDEGKSSGDNHLLRDPHYVEDNGPKYLVE
 ANTGTNDGQDGTSLNKAIFYGRSQSFTEKDKQLLIDTNKKHDSLANGALGIIELN
 NDTYTLKEMKPLIASNTVTDETERANVFKMNGRWYLFYDSSGSKMTINGISKDIYML
 GFSSNSLTGPKYPLANKTGLVNLNLDPTDFTFYSHFAVPQTNGKNVVTITSYITNRGM
 YSDHSSFPAPFLNLNKGTKTSVINSILQOQLTIDNY"
 3249..4793
 /gene="CAC1773"
 /note="sacC"
 3249..4793
 /gene="CAC1773"
 /codon_start=1
 /transl_table=11
 /product="Levanase"
 /protein_id="AAK79738.1"
 /db_xref="GI:15024743"
 /translation="MKRKLKLTLMIFTVILVLLSLYIQNSTIKSTTTKKQNY
 REVYHFTVPNKWKNDPQRPVFDGEHYHYLYLNKDYNGNGTEMRQATSKDLTWKDE
 GVCIPKTYNENGDIWGSFVYDAQNTAGFGKAIVAIIVTQPSASMDKOEQLWSTDR
 GRTEPKYSNQPILINPCDKDPDKPIIWDFFKNKWMVLAEGTKIGYESYNLKNMOH
 TGDEFTNNIGIVECPDIPFMOSNDGNYKWLGTSAKGVSGEPNTYAYWGVNQGKF
 IADISTPKWLDYGFDMYAAVTFESENQSKNLKRYALAMNNNDYANNPTPIQKRF
 MDSVREITKSKQDNTYSLSKRIENITTSIDQFOKISVGLKHLKVOGKAYOL
 DTDINWDAKNVGIRLRESLDKRRHDIGITECKYSVVRNAYTGNEDKSKKYVESRA
 PFDINNKVHLRIFVDKVSVEVFIDDDGKIYSNEVFRPDKGTILFSIKGAVESKNI
 TIKHID"
 4860..6146
 /gene="CAC1774"
 /note="sacB2"
 4860..6146
 /gene="CAC1774"
 /codon_start=1
 /transl_table=11
 /product="Levansucrase"
 /protein_id="AAK79739.1"
 /db_xref="GI:15024744"
 /translation="MKTRKTYKMISSLMVILAILTIPFLILRHNTGYTSIWSRQAO
 FKCTKENTAPINPNFKLTAPNLWVWDTWPLVKDGLAVVNGKYLIFALTASRNVGM
 NKRDHVGISYFCSTGDNVYKGLAYNVEDALGSRWAGSAILDENGWVQFYATG
 RKGFAVTFEORLVKTFKFSINVDKGGVHIITNSKHQVILEPDGVYVYTMQOAGPTIY
 SPDRPFEDQTKDYLIFEGNKGKIEKMPENIGDKLFRKNHIAPRGVEVNTIY
 GIVAAQNKDTREKLLPPLLEAVGQOQLERPOIVMKKNKYILFTLSHKFTYAQGLNG
 VGLYGFCCGNSLRNKLPLNGNLVITNPTNDPQYTSWYLVSGHDVLSFINHYHFG
 QLRYGTFAPTLQISLKGYSKIIRLGEVVTTPAH"
 6560..8305
 /gene="CAC1775"
 6560..8305
 /gene="CAC1775"
 /codon_start=1
 /transl_table=11
 /product="Predicted membrane protein"
 /protein_id="AAK79740.1"
 /db_xref="GI:15024745"
 /translation="MKIGIIAISMILLIGSISSVVRKASSNNKRLQNLTYKEFOKLK
 LKNSKSLVRDNDNNEKLLDLDKNNKVIPIFNKILKILPIIKSQDSSVSKYK
 GNSVASAGDKNNAAPVAKLDPIILNPESLNGEITNTQIAWISDSTDADGALSRA
 VGGFTSYILGVLDNKGFTVITNPAGDFDLYQVTDGSKGSLNIGYSIHVYPEDY
 QAFQGTFSQTTDVKTVNAVDFSKMDTAICLTGTSKOYMNKLKISDQSSQVQVNY
 DKWVYINKPSSNAGICNYITITPTVLSNSTYNSYRLMIGKNKNDVEAMLSGIQNTI
 ILDKVTNENKQFIQEYTPNKNQEWKFTADGSTVITLLNNYPERFORVDRPTENI
 FLDNPNNASTRSQCLNLYKAEKARLNTAGQDYLVITYSPSLIDQPTLNDIHL
 LVGDPLIYDSTTAYADNYITGPTTGYSSIANINVTNPTTAAVKDIYVRSNDVAR
 VSDFNKVMFPGESFWRDGGQCCIGIDVNVENGNTNNVQLMGTRFVFOQAGVDPI5

```

IIPGLHEDYVVEAGD"
8693..9346
/gene="CAC1776"
CDS
8693..9346
/gene="CAC1776"
/codon_start=1
/transl_table=11
/product="HAD superfamily hydrolase"
/protein_id="AAK79741.1"
/db_xref="GI:15024746"
/translation="MDRILFDYCYGLIDTKPVPDAISNIAKNILNSKDAVNIYVY
EDRLMYETIPYDKLIKQVLEYCDMELVTNVFSKSYDEIISVHKRLKPKPEVVESE
KIKKGGYALMSNVSHDIIDYNLDALGNVFNFLSELDHAYKPOLKPKFYVERLE
LKOKKHCHIAKGYWWDIVPCSKLGRKIWNRENKKNKCHPEYEEVHTLDEVINLI"
9598..10521
/gene="CAC1777"
CDS
9598..10521
/gene="CAC1777"
/codon_start=1
/transl_table=11
/product="Nudix (MUT-like) hydrolase"
/protein_id="AAK79742.1"
/db_xref="GI:15024747"
/translation="MRDMNKNRNLGTEKFEFLDRYPDYERPNTVDMLLFTVDDVP
IEQKDPKDKLILIKRGDHPYMGCAVPGGFVNI NEGLSSACYRELKKEETNVENVYF
EOLKTFGDDVNRDPRMRVIVSVAYMALADKLSITPKAGDDDDAKWFTVKRFTSSKGF
GIDKIDTYNILLSDDGETKIGILVTEKFEKNGVVTIKVPSYRLLEWSKEBELAFDHE
EIDCALERANKIEYPTAFSRPLKPYFTLREAOKVFEAILNLEKPLTRANFRFRKIKM
VWTEKEKTSRGPATCYTFNEDWQHTFLDE"
10543..11193
/gene="CAC1778"
CDS
10543..11193
/gene="CAC1778"
/codon_start=1
/transl_table=11
/product="Amidase from nicotinamide family"
/protein_id="AAK79743.1"
/db_xref="GI:15024748"
/translation="MCKFEKILNMSEIKGFVDLNDLKDKTATVIVDMVNGFVHE
GALSSRVEGIVDEIVRINKETLGNKIFFLDEHTNNSTEFKSYAKHCLGSLAEALI
PELNKALLDSNVIMPKNSVNGFHAPGKFLWLENEISOIENYIICCEYDVCVSNFA
NTLKTYPNOKMDKRIIPSNVAVTEFDGTHDGLMKIILSWEMQSGNIEIVDRL"
BASE COUNT 4466 a 1373 c 1945 g 3534 t
ORIGIN

Query Match 4.1%; Score 97.8; DB 1; Length 11318;
Best Local Similarity 49.2%; Pred. No. 1.3e-08;
Matches 656; Conservative 0; Mismatches 572; Indels 105; Gaps 11;

Qy 691 acagttccattctttaagcaagtgaaatacaataatcgctgcgcgtacacataaagat 750
Db 1862 AAGTGCGCTCAATTTAATGCATCAACATTAATAAATAATAGCTTCAGCAAGGGTTATGAT 1921

Qy 751 gcacaaactaatactatgaaccttttagatgtatggattcatggccagttccaagattg 810
Db 1922 AAGAAATGGTAATTTAGTAGA----TTTGGATGTATGGGTAGCTGGCCATTCGCAAAATGGT 1978

Qy 811 cggacagacaaagttgctaattggaatggctatcaacttgcacatcgcaaatgatgggaatt 870
Db 1979 GA----TGAACAGTGGGTAACTACCATGGGTATCATATTTCTTTCGCTTTGGCAGGTGAT 2035

Qy 871 cc-----aacacaaaatgatcatctatctcttataataaagatgtggtgataaagaa 927
Db 2036 CCTAAAAATCAAGATGATATCTTATTTATATGTTTTATCAAAAGATGGAGAAAATCC 2095

Qy 928 ttaagtcatggaagaatgtagtcccaatttttggctataattctaccgcggtt----- 981
Db 2096 ATTGATAGTTGGAGAATGCGAGTAAAGTGTTTAAGATAGTGTATAAATACGTTGCCAAT 2155

Qy 982 -----tcacaagaatggtcaggatcagctgttttgaacagtcgat 1020
Db 2156 GATCCTTATCTTAAAGTATCAAAACAGAAGATGGTCTGGTTCTGCAACTTTAAACATCTGAT 2215

```

```

Qy 1021 aactctatccaattattttata-----caagggtagacagctctgat 1062
Db 2216 GGTCAAGTTCTGTTGTTTACACAGATTTTCGGGAGTAGCAAAAGATGGTGAACAGAT 2275

Qy 1063 aacaataccaatcaataaaattgctagctactcttttatttaactgataataatgga 1122
Db 2276 GCTAGTAATCAAGTTATAACAACACTACTCAAGTGAATTTATCTCAACACGATCAAAATCA 2335

Qy 1123 aatglatcaactcgctcaggtacgaataactatattgtatttgaagggtgagctattac 1182
Db 2336 ATTAATATTGATGATGTAAGTATGATATAAATCTGTTTTTACGCGGAGAAATGGGCAATTC 2395

Qy 1183 taccaaaacttatgatcaatggaaagctactaacaagg-----tgcgcgataattgca 1236
Db 2396 TATCAGAAGCTTCAGCAATTTATAGATGAAGTAAATGGAGTCTGGCGATTAATCATACC 2455

Qy 1237 atgctgatgctcgtatgaattgaagatgtaattgtaactgttctgttttgaagca 1296
Db 2456 TTGAGAGATCTCTCATTTATGTGGAAGA---TAATGGCCGCAAAATATCTTTGTTTCGAGGCT 2512

Qy 1297 agtactggtttggaaaaat---tatcaagcgagagaccacaaattataactgggttaattat 1353
Db 2513 AACACAGGACAAACGATGGTTACCAGGAGATACTTCTCTACTCAATAAAGCTTTTTAT 2572

Qy 1354 ggcggagatgacgcatttaataatcaagagcttatttagaattcttccaatgatgatt 1413
Db 2573 GGTAGAAGTCAAGTTCTTTAAACAGAAAAGATCAATTTAATTGATACTAATAA 2632

Qy 1414 aagagtcgggcaactgggctaactcagctatcggtatcctcaactaaataaagacgaa 1473
Db 2633 AAGCATGATGCTTCTTTAGCTAATGGGCACCTAGGCATTTATGAATTAATAATGACTA- 2691

Qy 1474 aagaatcctcaaggtggcagagttatactcaccatttaattctgcacaaatgtaagcgat 1533
Db 2692 -----TACATTAAAAAAGAAATGAACCATTAATTGTCATCCAACTACTGTACAGAT 2743

Qy 1534 gaaattgacgacaaaatgtagttaaattagtgtaataataattacttatttccgcctacc 1593
Db 2744 GAAATTGAACGTCGTAATGTATTAAAAATGAATGTTAGTGTGATGCTACTATTTACAGATTCA 2803

Qy 1594 cgtttaaatcagaggaagtaaatgatgcttggaatgtaactgtaataatgcgcttggtgat 1653
Db 2804 AGAGGATCTAAATGACTATTAAATGAATAAGTTCAAAAGATATATA----- 2850

Qy 1654 aatgttgaatggctcgggatatgttgcgtatgcttaactggaatcttataaagcattaaat 1713
Db 2851 -----CATGCTAGGTTTATAGTTCAAAATTCCTTGACTGGTCCATACAAACCATTTGAAT 2902

Qy 1714 gattcggagtagtcttgactgctctctctcctgcacaaactggcgagacgaactattca 1773
Db 2903 AAACTGGCCCTTGTAATAAATTTAAACCTTGATCTCTACTGATCTAACATTTTACGTACTCT 2962

Qy 1774 tattatgctgtcccgcttgcgcgaaaagatgaccgaagtattgatttacttcataatgat 1833
Db 2963 CATTTGCTGTCTCCCAACAACTAATGGTAAAAACGTAGTAATAACAAAGTTAC----- 3013

Qy 1834 aatagaaatggagtagcgggttaaagggaatgattcaacttgggcaccgagttcttacta 1893
Db 3014 ATAACAAATAGAGGAATGTATTTCAGATCATCTACTGTAGTTTGGCACCAGTTTCTACTT 3073

Qy 1894 caaattaacccgggataacacaaactactgttttagctaaaaatgactaaatcaagggttgg 1953
Db 3074 AATATTAAAGGAACCAACAAACATCTGTTATATCAATAGCATTTCTTCAGCAAGGCAACTT 3133

Qy 1954 atttgggatgatt 1966
Db 3134 ACTATTGATAATT 3146

```

RESULT 8
AC007926/c
LOCUS

AC007926

155204 bp

DNA

linear

HTG 17-JUL-2001


```

/note="pot. rRNA binding site (sacB)"
464..550
464..1885
/codon_start=1
/transl_table=11
/product="levansucrase"
/protein_id="CAA26513.1"
/db_xref="GI:732568"
/db_xref="SWISS-PROT:P05655"
/translation="MNKKFAKQATVLTFTTALLAGGATQAFKFNOKPKYKTYGIS
HTRHMDLQIPQOKNEKYQVPEFDSSTIKNISKAGLVDWSDPLOWADGTVANVHG
YHIFALAGDPKNADDTSYMFYQKVGSTISDWKNAKGRVFKDSKDPANDSLIKDOT
QWMSGATFSDGKIRLFTVDFSGHKYKQTLTAAQVNVASDSSLNNGVEDYKSF
DGDGTYNQVOQFIDEGNYSNGDNLTRDPHYVEDKHKHLYLFEANTGDEYGOEES
LFNKAYYKSTSFQESQKLLQSKKTAELANGALGMIELNDYTLKKYMKPLIAS
NYVTEIERANVFNRKNGKWLFTDSRGSMTIDGITSNDIYMLGYVSNLSLTPKPLN
KTVGLVKMDLPNDVFTYSHFAVPOAKGNVVIITSYMTNRGFYADKOSTAPSPFLN
IKGKTSVYVKDSILEQGLTVNK"
551..1882
/mat_peptide
/product="mature levansucrase (aa 1-444)"
1892..1904
/repeat_unit
/note="inverted repeat B"
1892..1928
/misc_feature
/note="region of dyad symmetry"
1916..1928
/repeat_unit
/note="inverted repeat B"
BASE COUNT 720 a 422 c 370 g 495 t
ORIGIN

```

Query Match 3.4%; Score 80; DB 1; Length 2007;
Best Local Similarity 47.7%; Pred. No. 3.3e-05;
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

```

Qy 983 cacaagaatggcagatcagctgtttgaacagtgataactctatccaaattatttata 1042
Db 942 CACAAGAATGGTCAGGTCAGCCACATTTACATCTGACGGAAAAATCGGTTATTCTACA 1001
Qy 1043 caagggtagacagctctgatacaataccaataccaataaattgctagcgaatgactattgtat 1102
Db 1002 C-----TGATTCTCCGGTAAACATTACGGCAACAACACTGCACTGCACAAAGTTA 1055
Qy 1103 attaacatgataataatggaatgtatcactcgtcaggtacgaatgactattgtat 1162
Db 1056 AGTATACATGATCAGACAGCTCTTTGAACATCAACGGGTAGAGGATTTATAATCAATCT 1115
Qy 1163 ttgaaggtgagctgtattactaccaaaacttatgatcaatgaa-----aagctactaacaag 1219
Db 1116 TTGACGGTGACGGAACAAACGTATCAAAATGTACAGCATTTTCATCGATGAAGGCAACTACA 1175
Qy 1220 gtgcgataaatttgcaatgcgtgatgctcatgtaattgaagatggtaattggtgatcgg 1279
Db 1176 GCTCAGGCGACAACCATACGCTGAGAGATCTCTACTAGTGTAGAGATAAAGGCCACAAAT 1235
Qy 1280 acctgtttttgaagcaagtaactgttttgaataattcaaggcgaggaacaaattata 1339
Db 1236 ACTTAGTATTTGAAGCAACACTGGAAGTGAAGATGGCTACCAAGCGCAAGAAATCTTAT 1295
Qy 1340 actggttaaatattggcggagatgacgcatttaataatacaagacttattagaattcttt 1399
Db 1296 TTAACAAAGCATACTATATGCGCAAAAGCACATCTTCTCCGTCAAGAAGTCAAAAACCTC 1355
Qy 1400 ccaatgatgataaagtagtggcaacttggcctaataatgacgctatcggtatctctcaaac 1459
Db 1356 TGCAAAACGATATAAAAACCGACGCTGAGTTAGCAAAACGGCGCTCTCGGTATGATTCAGC 1415
Qy 1460 taaataaggcagcaaaagaactcctaagtgccagattataactcacaattattcttcgac 1519
Db 1416 TAAACGATGATTACACTGTAAGAAAGATGATGAA-----ACCGCTGATTGCATCTA 1466
Qy 1520 caatggttaagcagtaaatgagcagcaaaatgtagttaaattaggtaataataattact 1579
Db 1467 ACACAGTAACAGATGAATTTGAACGCGCAACGCTCTTTAAATGCACGCGCAATGGTACC 1526

```

```

Qy 1580 tattccgcctaccggttttaaatcagaggaagtaaatgatgatccttgatgaatgctaat 1639
Db 1527 TGTTCACCTACTCCCGCGGATCAAAAATGACGATTGACGGCATTTACGCTAACGATATTT 1586
Qy 1640 atgcgctgtgataaatgttgcgaatgttcgagatatgttgcgtatgctgaactgaatctt 1699
Db 1587 A-----CATGCTTGGTTATGTTCTTAATTTCTTAACTTGCCCAT 1625
Qy 1700 ataaaccattaaatgattcgtgagtagtgctgactgcttctctctcctgcaaaactgacga 1759
Db 1626 ACAAGCCGCTGAACAAAACCTGGCTTGTGTAAAAATGGATCTTGATCTCAACGATGTA 1685
Qy 1760 cagcaacttattcattatgctgccccgttgcggaaagatgacccaagtattagttta 1819
Db 1686 CTTTACTTACTCACATCTCGCTGTACC---TCAAGCGAAAGGAACAATGTCGTGATTA 1742
Qy 1820 ctctatatgactaatagaaatggtagcggtggtgaaaggaatggatcaacttgggcac 1879
Db 1743 CAAGCTATATGACAAACAGAGGATTCTACGCAGACAAA-----CAATCAAGTTTGGC 1796
Qy 1880 cgagttcttactacaataaccgggataacacaaactactgttttagcttaaaatgacta 1939
Db 1797 CAAGCTTCTGCTGACATCAAAAGCAAGAAACATCTGTTGTCAAGACAGACATCCTTTG 1856
Qy 1940 atcaaggggatt 1951
Db 1857 AACAGGACAAT 1868

```

```

RESULT 10
SYNPRLC/c
LOCUS
DEFINITION SYNPRLC 5941 bp DNA circular SYN 27-APR-1993
double recombinants in gram-negative bacteria.
ACCESSION L05083
VERSION L05083.1 GI:209132
KEYWORDS Cloning vector.
SOURCE Cloning vector.
ORGANISM unidentified cloning vector
artificial sequence; vectors.
REFERENCE 1 (bases 1 to 5941)
AUTHORS Cai, Y.
TITLE Molecular genetic approaches towards the understanding of
heterocyst differentiation and pattern formation in the
cynobacterium Anabaena sp
JOURNAL Unpublished (1992)
REFERENCE 2 (sites)
AUTHORS Black, T.A., Cai, Y. and Wolk, C.P.
TITLE Spatial expression and autoregulation of hetR, a gene involved in
the control of heterocyst development in Anabaena
JOURNAL Unpublished (1992)
FEATURES
Location/Qualifiers
source 1..5941
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
complement(473..1894)
/note="sacB from B.subtilis; ORF"
/transl_table=11
/protein_id="AAA72307.1"
/db_xref="GI:209133"
/translation="MNKKFAKQATVLTFTTALLAGGATQAFKFNOKPKYKTYGIS
HTRHMDLQIPQOKNEKYQVPEFDSSTIKNISKAGLVDWSDPLOWADGTVANVHG
YHIFALAGDPKNADDTSYMFYQKVGSTISDWKNAKGRVFKDSKDPANDSLIKDOT
QWMSGATFSDGKIRLFTVDFSGHKYKQTLTAAQVNVASDSSLNNGVEDYKSF
DGDGTYNQVOQFIDEGNYSNGDNLTRDPHYVEDKHKHLYLFEANTGDEYGOEES
LFNKAYYKSTSFQESQKLLQSKKTAELANGALGMIELNDYTLKKYMKPLIAS
NYVTEIERANVFNRKNGKWLFTDSRGSMTIDGITSNDIYMLGYVSNLSLTPKPLN
KTVGLVKMDLPNDVFTYSHFAVPOAKGNVVIITSYMTNRGFYADKOSTAPSPFLN
IKGKTSVYVKDSILEQGLTVNK"
2749..3543
/note="kanamycin and neomycin resistance determinant"
CDS

```


/transl_table=11
/protein_id="AA072304.1"
/db_xref="GI:209128"
/translation="MEKIKGYTTVDISOWHKEHFEAFQSVAGCTYNQTVOLDITAF
LKVKKKKHKKYPAFTHILARLNNAHPEFMAKMDGELVWDSVHCYTVFTHQETETP
SSLKSEVHDDPFRFLHIYSODVACYGENLAIFPKGETENFFVSANPWSVTSFEDLNV
ANWNEFAPVFTMGKITYTQDKVLMPLAQVHHAVCDGFHVHRLMELQYCDWQGG
A"

misc_feature 4252..5221
/note="multiple cloning sites"
misc_feature 5222..6347
/note="oriV and oriT genes from Plasmid pMB1"

BASE COUNT 1649 a 1305 c 1404 g 1989 t
ORIGIN

Query Match 3.4%; Score 80; DB 12; Length 6347;
Best Local Similarity 47.7%; Pred. No. 2.8e-05;
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

Qy 983 cacaagaatggtcaggatcgctgttttgaaacagtgataactctatcccaattattttata 1042
|||||
Db 1416 CACAAGAATGCTCAGGTTCAGCCACATTTACATCTGACGGAAAAATCCGTTTATTCTACA 1357
Qy 1043 caagggtagacagctgtgataacaataccaatacatcaaaaaattgctagcgtactcttt 1102
|||||
Db 1356 C-----TGATTCTCGGTAACATTTACCGGCAAAACAACTGACAACTGCACAAAGTTA 1303
Qy 1103 atttaactgataaatggaattgatactcgtcgtcaggtacgaaatgactattgtat 1162
|||||
Db 1302 ACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTAGAGGATTATAATCAATCT 1243
Qy 1163 ttgaagtgatgcttattactacaaacttatgatcaatgga---aagctactaaacaaag 1219
|||||
Db 1242 TTGACGGTGCAGGAAAAACGATATCAAAATGATACAGCAGTTTCATCGATGAAGGCAACTACA 1183
Qy 1220 gtgcgataatttgaatcggtgctgctcatgtaattgaagatgtaataatggtgatcggt 1279
|||||
Db 1182 GCTCAGCGCACACCATACCTCAGAGATCCTCCTACGTACGTAGAAGATAAGGCCACAAAT 1123
Qy 1280 acctgttttgaagcaagtaactggtttggaaaattatcaaggcgagggaccaattata 1339
|||||
Db 1122 ACTTAGTATTGTAAGCAAAACACTGGAAGTGAAGTGCCTACCAAGGCGGAAGATCTTTAT 1063
Qy 1340 actggttaattatggtcgagatgacgcatttaataatcaagcgtatttagaattcttt 1399
|||||
Db 1062 TTAACAAGCATACTATGCAAAAGCACATCATCTTCCTCGTCAAGAAAGTCAAAAACCTTC 1003
Qy 1400 ccaatgatgatattaagagtcgggcaacttgggctaatgcagctatcggtatctctcaaac 1459
|||||
Db 1002 TGCAAAAGCGATAAAAAACGACGCGTGAGTTAGCAAAACGCGCTCTCGGTATGATTGAGC 943
Qy 1460 taataagcagcaagaataacttaagtggtgagagtgataactacacattaatcttcgcaac 1519
|||||
Db 942 TAAACGATGATACACACTGAAAAAAGATGATGAA-----ACGCGCTGATTGCTATCA 892
Qy 1520 caatggtgaagcagatgaattgagcgaccataatgtagttaattaggtataataattact 1579
|||||
Db 891 ACACAGTAACAGATGAATTAAGCGCGCAAGCTCTTTAAATGAACGCGCAATGGTATACC 832
Qy 1580 tatttgcgctaccgctttaaactcagggaagtaatgatgatgcttggatgtaattgcttaatt 1639
|||||
Db 831 TGTTCACTGACCTCCCGCGGATCAAAAATGACGATTGACGGCATTAAGTCTAACGATATTT 772
Qy 1640 atgcgcttggtgataatgttgcaatggtcgatattgctctgatagctcaactggtattct 1699
|||||
Db 771 A-----CATGCTGGTATGTTTCTTAATCTTTAACTGCCCAT 733
Qy 1700 ataaagccataaatgattctgtagtagtcttgcactgcttcttgcgcaactggcgga 1759
|||||
Db 732 ACAAGCGCTGAACAAACTGGCCTGTGTTAAAAATGGATCTTGATCTCTAACGATGTA 673
Qy 1760 cagcaacttattcattatgctgtccccggttgcgggaaagatgaccaaagtattagttta 1819

Db 672 CCTTTACTTACTACACTTCGCTGTACC---TCAGGGAAGGAAACAATGTCTGATTA 616
Qy 1820 cttcatatgactaatagaaatggagtagcggtggtgaaagaaatggatcaacttgggcac 1879
Db 615 CAAGCTATATGACAAACAGAGGATTCTACGACAGACAAA-----CAATCAACGTTTGGCG 562
Qy 1880 cgagtttctactacaaattacccgggataaacacaaactactgttttagctaaataacta 1939
Db 561 CAAGCTTCTCTGCAACATCAAAAGGCAAGAAACATCTGTTGTCAAAGACAGCATCTCTG 502
Qy 1940 atcaaggggatt 1951
Db 501 AACAAAGGACAAAT 490
RESULT 12
SYNPRUB/c
LOCUS
DEFINITION SYNPRUB 6798 bp DNA circular SYN 27-APR-1993
Cloning vector (pRL277) for sacB-mediated positive selection for
double recombinants in gram-negative bacteria.
ACCESSION L05082
VERSION L05082.1 GI:209129
KEYWORDS Cloning vector.
SOURCE Cloning vector DNA.
ORGANISM unidentified cloning vector
artificial sequence; vectors.
REFERENCE 1 (bases 1 to 6798)
Cai,Y.
AUTHORS Molecular genetic approaches towards the understanding of
heterocyst differentiation and pattern formation in the
cynobacterium Anabaena sp
Unpublished (1992)
JOURNAL 2 (sites)
AUTHORS Black,T.A., Cai,Y. and Wolk,C.P.
TITLE Spatial expression and autoregulation of hetR, a gene involved in
the control of heterocyst development in Anabaena
Unpublished (1992)
JOURNAL Location/Qualifiers
FEATURES
source 1..6798
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
complement(473..1894)
/note="sacB gene from B.subtilis; ORF"
/codon_start=1
/transl_table=11
/protein_id="AA072305.1"
/db_xref="GI:209130"
/translation="MNKKFAKQATVLTFTTALLAGGATQAFKQKPKYKTYGIS
HITRDMQIQPEQKNEKYQPEFSDSTIKNISAKGLEWDSWPLQNDVANYHG
YHIFALAGDPKNADDTSYMFYQKVGSTSIDSWKNAGRVFKSDKFDANDSILKDOT
QEWGSGATFTSDGKIRLFTYDFSGKHGKQTLTAQVNVASDSLSNLINGVEDYKSF
DGDGKTQNVQOQFIDEGNYSNGDNHTRDPHYVBDKHGKYLVEFANCTGDEGYGSES
LFNKAYYKSTSEFROSKYLQSDKKRTAELANGALMIELNDYTLKKVKKELIAS
NTVDELERANVFKMGKYLFTDSRSKMTIDGITSNDIYMLGTVSNLSLPGPKPLN
KTGLVLMKMDLPNDVFTTYSHFVAPQAGNNVITSYMTNRFYADKQSTFAPSFLN
IKGKTSVVKDSILEQGLTVNK"
complement(2980..3771)
/note="streptomycin and spectinomycin resistance
determinant (aadA)"
/codon_start=1
/transl_table=11
/protein_id="AA072306.1"
/db_xref="GI:209131"
/translation="MREAVIAEVSQTSVGVVIERHLEPTLLAVHLYGSAVDGGLKP
HSDTLLVTVTRLEDTTRALINDLETASPCGESILRAVEVTIVHDDIIPWRYP
AKRELQFGWQRNDILAGIFEPAITDIDLAILLKAREHSVALGPAARELFDVPEQ
DLFALNETLTANSPPDWAGDERNVLTLSRIWYSAVTQKIAKPAKVAADWAMERLPA
QYQPVILEARQAYLQGBEDRLASRADQLEEFVHYVKGKTEIKVVGK"
4241..5504
/note="multiple cloning sites"
5667..6798
/note="oriV and oriT genes from Plasmid pMB1"
misc_feature
misc_feature

```
BASE COUNT      1653 a   1540 c   1637 g   1968 t
ORIGIN

Query Match      3.4%; Score 80; DB 12; Length 6798;
Best Local Similarity 47.7%; Pred. No. 2.8e-05;
Matches 464; Conservative 0; Mismatches 460; Indels 48; Gaps 6;

QY 983 cacaagaatggtcaggatcagctgttttgaacagtgataaactctatcccaattatttata 1042
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 CACAAGAATGTCAGGTTTCAGCCACATTTACATCTCAGCGAAAAATCCGTTTATTCTACA 1357
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1043 caagggtagacagctgatacaaatcaccaatcatcaaaaaattgctagcgtactcttt 1102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1356 C-----TGATTTCTCGGTAACATTTACGGCAAAACAACTGACAACTGCAACAAGTTA 1303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1103 atttaactgataatgaaatgtatcactgcctcaggtacgaaatgactatattgtat 1162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1302 ACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTAGAGGATTTAAATCAATCT 1243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1163 ttgaaggtgatgctattactaccaaacttatgatcaatgga---aagctactaacaag 1219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1242 TTGACGGTGACGGAAACAGTATCAAAATGTACAGCATCTCATCGATGAAGGCAACTACA 1183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1220 gtgcccgaataattgcaatgcgtgatgctcatgtaattgaagatggttaatgctgct 1279
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1182 GCTCAGCGCACACCATACCGCTGAGAGATCCTCACTACGTAGAGATAAAGGCCACAAAT 1123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1280 acctgtttttgaagaagtaactgtgttggaaaaatatcaagcgagaccataatttata 1339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1122 ACTTAGTATTTGAAGCAAACTGGAACCTGAAGATGGCTTACCAAGCGGAAGAATCTTTAT 1063
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1340 actggttaattatgcgagagatgcgcaatttaatacaagagcttattagaattcttt 1399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1062 TTAACAAAGCATACATGSCAAACACATCATCTTCCGTCAAGAAAGTCAAAAACATTC 1003
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1400 ccaatgatgataatgaagtcgtggcaacttgggttaatgcagctatcgttatccctcaaac 1459
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1002 TGCAAAGCGGATAAAAGCGACGGCTGTAGTGTAGCAACGGCGCTCTCGTGTATGATTGAGC 943
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1460 taataagagcaaaaagaatcctcaaggtgagaggttatactcaccataattttgcac 1519
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 942 TAAACGATGATTACACATGTAAGAAAGTGATGAA-----ACCGGTGATTGTCATCTA 892
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1520 caatggttaagcgatgaattgagcgaccataatgtagttaattaggtaataaattact 1579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 891 ACACAGTAACAGATGAATTTGAACGCGCAACGCTCTTTAAATGAACGGCAATGGTACC 832
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1580 tatttgcgcgtcaccgctttaaactcaggaagtaatgatgatgcttggatgaatgctaatt 1639
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 831 TGTTCACATGACTCCCGGGATCAAAAATGACGATTTGACGGCATTAACGATATTT 772
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1640 atgcccgttggatgaatgttgcgaatggttgcgtgatgttgcgtgaagtaactgacttt 1699
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 771 A-----CATGCTGTGTTATGTTTCTTAATCTTAACTGGCCCAT 733
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1700 ataagcattaaatgattctgagtagcttcttgactgctctctgttcctgcgaactggcgga 1759
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 732 ACAGCGCTGAACAAACATGGCCCTGTGTAAATAATGGATCTTGTATCTCAACGATGTA 673
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1760 cagcaacttattcatattatgtctcccgcttgcggaagaagatgacaaagtattagttta 1819
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 672 CTTTACTTACTACACTTCGCTGTACC---TCNAGCGAAGGAAACAAATGTGCTGATTA 616
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1820 ctccatataactaataagaaatggaatgagtagcgggttaaaggaatggattccaacttgggcac 1879
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 615 CAAGCTATATGACAAACAGAGGATTTCTTACGCAGACAAA-----CAATCAACCTTTGGCG 562
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1880 cgaatttctactacaattaaaccgggataaacaactactgttttagctaaatgacta 1939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 561 CAAGCTTCTGCTGAACATCAAGGCAAGAAACATCTGTGTGTCAAAGACAGATCCTTG 502
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1940 atcaaggggatt 1951
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 01:51:26 ; Search time 7643.51 Seconds
(without alignments)
8182.745 Million cell updates/sec

Title: US-09-995-587A-10
Perfect score: 4634
Sequence: 1 gttacaagacaaaatttt.....gttactgcttgcacaaagctt 4634

Scoring table: IDENTITY NUC
Gap 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_estl.*
10: gb_est2.*
11: gb_hic.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	2.3	641	12	AQ946120 Sheared D
2	97	2.1	700	12	AQ940248 Sheared D
3	92.8	2.0	687	12	BH401326 Sheared D
4	91.4	2.0	806	12	AL281759 Tetraodon
5	89.8	1.9	832	12	BH391984
6	89.4	1.9	729	12	AQ945618
7	84.6	1.8	469	9	AU076358
8	82.6	1.8	401	12	AQ945618 Sheared D
9	81.2	1.8	621	12	AQ945618 Sheared D
10	81	1.7	500	9	AU086536
11	81	1.7	773	12	BH391959
12	80.2	1.7	520	12	AQ945618 Sheared D
13	79.4	1.7	541	12	AL295309 Tetraodon
14	79.4	1.7	639	12	AQ945618 Sheared D
15	79.2	1.7	560	12	AQ945618 Sheared D
16	78.2	1.7	713	12	AZ217550
17	78	1.7	879	12	CNS01JUG

18	77.4	1.7	252	9	AU052594
19	77.4	1.7	540	9	AU037631
20	76.4	1.6	584	10	BM163760
21	76.4	1.6	1007	12	CNS06X9S
22	76.2	1.6	660	12	AZ212542
23	76	1.6	498	10	BM276013
24	75.4	1.6	519	10	BI814949
25	75.4	1.6	579	9	AU038989
26	75.4	1.6	781	12	AQ952743
27	75.2	1.6	1025	12	CNS01472
28	75	1.6	653	12	AQ311782
29	74.4	1.6	522	10	BI815912
30	74.4	1.6	1101	12	CNS00396
31	74.2	1.6	305	9	AU038805
32	73.8	1.6	471	12	BH391050
33	73.8	1.6	687	12	AG145469
34	73.6	1.6	660	9	AU038978
35	73.2	1.6	579	9	AU038185
36	73.2	1.6	743	12	AZ196612
37	73	1.6	553	12	BH368964
38	72.4	1.6	522	12	BH369425
39	72.4	1.6	581	12	AQ658242
40	72.4	1.6	950	9	AU175513
41	72.2	1.6	410	9	AU037570
42	72.2	1.6	489	12	BH444866
43	71.6	1.5	907	12	CNS02134
44	71.2	1.5	500	9	AU087262
45	71.2	1.5	571	10	BM274158

ALIGNMENTS

RESULT 1
LOCUS AQ946120/c
DEFINITION Sheared DNA-46J23.TR Sheared DNA Trypanosoma brucei genomic clone
ACCESSION AQ946120
VERSION AQ946120.1 GI:6769385
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE 1 (bases 1 to 641)
AUTHORS El-Sayed, N., Zhao, S., Gill, S., Suh, E., Malek, J., Fujii, C., Gerard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J., Fraser, C., and Adams, M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: Sheared DNA-46J23.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mbd/tbdb/.
Seq primer: M13-Reverse
Class: shotgun.

FEATURES

Location/Qualifiers
1..641
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-46J23"
/clone_lib="Sheared DNA"

DB 7AC

Qy 1448 actaacaaactcgaagaaatcgtgttgaatgaagcgaataatacctgaaat 150/
Db 746 caccacattatataatataatataatataatataatataatataatataat 687

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tldb/mdb/tbdb/>.

Seq primer: M13-Forward

Class: Shotgun.

Location/Qualifiers

FEATURES

source

1. 729

/organism="Trypanosoma brucei"

/strain="TREU927/4 GUTat 10.1"

/db_xref="taxon:5691"

/clone="Sheared DNA-49K6"

/clone_lib="Sheared DNA"

/note="Vector: pUC18; Site_1: SmaI; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically

sheared to give a tight size distribution (approx 2 kb).

The v + i method used for the library construction is

described in detail in Smith, H.O. and Venter, J.C.

(Making small insert libraries for whole genome shotgun

sequencing projects. In Genome Sequencing: A Practical

Approach, eds. M. Vaudin and B. Borell, Oxford University

Press, 1999).

360 a 58 c 58 g 253 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 1.9%; Score 89.4; DB 12; Length 729;

Matches 198; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1359 aagctaagtgcacaaattggcgtagtggaataattattagtgctaacagtaatacaacca 1418

Db 349 AAGTTATTGTGTATATAGCAGTAATAATAATAATAATAATAATAATAATAATAATA 408

QY 1419 gggcgcgacaaatcaagaatcatctgctactcaacaatactgaaatgctgttgtaag 1478

Db 409 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 468

QY 1479 aaagcaaaaataactaactgctgaaatgctgttgtaataagaaacaaaatactcaaca 1538

Db 469 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 528

QY 1539 atactgaaatgctgttgtaagaaacaaaatactcaacaacagaaaacgataata 1598

Db 529 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 588

QY 1599 gtcaattaaagttaactaataatgaacaccatcgccgctactcaagcaaaactggaaga 1658

Db 589 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 648

QY 1659 agctaaatctcaagctgaaggctgttcaaaatgccagattgacgcggtgatttaa 1718

Db 649 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 708

QY 1719 cagatgatacaaatgaatga 1737

Db 709 ATAATAATAATAATAATAATA 727

RESULT

AU076358

LOCUS

DEFINITION AU076358 Dictyostelium discoideum 469 bp mRNA linear EST 27-JUL-1999

dictyostelium cDNA clone SSA642, mRNA sequence.

ACCESSION AU076358

VERSION AU076358.1 GI:5607256

KEYWORDS Dictyostelium discoideum.

SOURCE Dictyostelium discoideum

ORGANISM Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

REFERENCE 1 (bases 1 to 469)

AUTHORS

TITLE

JOURNAL

COMMENT

Urushihara,H.

Developmental CDNA in Dictyostelium discoideum (1999)

Unpublished (1999)

Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan

Email: d402hu@sakura.cc.tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

Location/Qualifiers

FEATURES

source

1. 469

/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="SSA642"

/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"

/dev_stage="slug"

253 a 30 c 26 g 157 t 3 others

BASE COUNT

ORIGIN

Query Match 1.8%; Score 84.6; DB 9; Length 469;

Best Local Similarity 53.2%; Pred. No. 2.7e-05;

Matches 177; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 1429 aaatcaagaatcatctgctacttaacaaatactgaaatgctgttgtaataatgaaagcaaaaa 1488

Db 5 AAATCAAGAATCAGTTCTCCAAGACATTATAATAATAATAATAATAATAATAATAATA 64

QY 1489 tactaacaatactgaaaaatgctgttgaatgaaacaaataactaataactgaaaa 1548

Db 65 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 124

QY 1549 tgcgtgttgaatgaaacaaataactaacaacagaaacgataatagtcacattcaa 1608

Db 125 TATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 184

QY 1609 gttactaataatgaacaaacctcagcgcgtactcaagcaaaccttgaaagcgaatcc 1668

Db 185 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 244

QY 1669 tcaagctgctaaaggctgttcaaaatgccagattgacgcggtagtttaacagatgaca 1728

Db 245 TAANAATAATAANAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 304

QY 1729 aattaagaattaaataagattacacttctctaa 1761

Db 305 TAATAATAATGTGAATAATGATAATGTAAATAA 337

RESULT

AU0639568/c

LOCUS

DEFINITION 927P1-4D3.TP 927P1 Trypanosoma brucei genomic clone 927P1-4D3, DNA

sequence.

ACCESSION AU0639568

VERSION AU0639568.1 GI:5116278

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE 1 (bases 1 to 401)

AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,

Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,

Fraser,C. and Adams,M.

Determination of clone end sequences from Trypanosoma brucei TREU

927/4 pl library

Unpublished (1999)

Other_GSSs: 927P1-4D3.TV

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@tigr.org
 For clone/filter availability, please contact Sara Melville
 (sml60@mol.bio.cam.ac.uk). Pl end sequences search page:
 http://www.tigr.org/tdb/mdb/tbdb/
 Seq primer: SP6
 Class: Pl ends

FEATURES

Location/Qualifiers
 1. 401
 /organism="Trypanosoma brucei"
 /strain="TREU927/4"
 /db_xref="taxon:5691"
 /clone="927Pl-4D3"
 /clone_lib="927Pl"
 /note="Vector: pAD10SacBII; Site_1: Bam HI; Constructed by Sara Melville, University of Cambridge, UK and Nancy Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was isolated from Trypanosoma brucei (stock TREU927/4) and partially digested with Sau 3AI. DNA fragments were cloned into the Bam HI site of pAD10SacBII vector (Genbank accession U09128). The average insert size is 65 Kb. Coverage: approx 4.4 X the haploid non-minichromosomal genome"
 130 a 39 c 7 g 225 t

BASE COUNT

ORIGIN

Query Match 1.8%; Score 82.6; DB 12; Length 401;
 Best Local Similarity 55.4%; Pred. No. 6.1e-05;
 Matches 160; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
 QY 1345 taatactgaacgctcaagctaatggcgaattggcgtagatggaaaattatttagtgcata 1404
 Db 395 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 336
 QY 1405 cagtaatacaaccagtggtgcgaataatcaagatcatctgctactcaacaactgaaaa 1464
 Db 335 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 276
 QY 1465 tgcgtgtgtaatagaagcaaaataactaacaatactgaaaatgctgttgaatgaaa 1524
 Db 275 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 216
 QY 1525 caaaatactacaatactgaaaatgctgttgaatgaaaatacaatactacaacac 1584
 Db 215 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 156
 QY 1585 agaaaacgataatagtcgaattaaagtaactaataatgaacaaccatca 1633
 Db 155 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 107

RESULT 9

AQ941683 621 bp DNA linear GSS 27-JAN-2000
 LOCUS Sheared DNA-43D2.TR Sheared DNA Trypanosoma brucei genomic clone
 DEFINITION Sheared DNA-43D2, DNA sequence.
 ACCESSION AQ941683
 VERSION AQ941683.1 GI:6764948
 SOURCE GSS.
 ORGANISM Trypanosoma brucei.
 Trypanosoma brucei
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.
 1 (bases 1 to 621)
 El-Sayed,N., Zhao,S., Gill,H., Zhao,H., Gyll,S., Suh,E., Malek,J., Fujii,C.,
 Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
 Fraser,C. and Adams,M.
 TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
 10.1 sheared DNA library
 JOURNAL Unpublished (1999)

COMMENT

Other_GSSs: Sheared DNA-43D2.TF
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through ATCC. Sheared DNA end sequences search page:
 http://www.tigr.org/tdb/mdb/tbdb/
 Seq primer: M13-Reverse
 Class: Shotgun.

FEATURES

Location/Qualifiers
 1. 621
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-43D2"
 /clone_lib="Sheared DNA"
 /note="Vector: pUC18; Site_1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds M. Vaudin and B. Barrell, Oxford University
 Press, 1999).
 318 a 15 c 78 g 210 t

BASE COUNT

ORIGIN
 Query Match 1.8%; Score 81.2; DB 12; Length 621;
 Best Local Similarity 53.9%; Pred. No. 0.00011;
 Matches 167; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 1314 aaagtaacaattacacgcggtgtgctgaagttaactactgaacgctcaagctaatggcaca 1373
 Db 299 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 358
 QY 1374 ttgcccgtagatggaaaattatttagctcaacagtaacacacccagtgctgcacaaatc 1433
 Db 359 ATGATAATAATAATAATGATGAGGAGTGTGTGAGTGTGTATATACGAATATTATAA 418
 QY 1434 aagaatcatctgctactacaactactgaaatgctgttgaatgaaagcaaaataacta 1493
 Db 419 AAACGAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 478
 QY 1494 acaatactgaaaatgctgttgaatgaaacaaaataactaacaatactgaaaatgctg 1553
 Db 479 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 538
 QY 1554 ttgttaatagaacaaaataactaacaacacagaaaataataatagctcaattcaagttaa 1613
 Db 539 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 598
 QY 1614 ctaataatga 1623
 Db 599 ATAATAATAA 608

RESULT 10

AU086536 500 bp mRNA linear EST 27-JAN-2001
 LOCUS AU086536
 DEFINITION SUGANO Malaria cDNA library Plasmodium falciparum cDNA
 clone xPfn2783, mRNA sequence.
 ACCESSION AU086536
 VERSION AU086536.1 GI:12388677
 KEYWORDS EST.

SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 500)
AUTHORS Watanabe, J., Sasaki, M., Suzuki, Y., and Sugano, S.
TITLE FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum
JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE 20574754
COMMENT Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanab@manage.ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES
source Location/Qualifiers
1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone="XPf2783"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 278 a 43 c 34 g 142 t 3 others
ORIGIN

Query Match 1.7%; Score 81; DB 9; Length 500;
Best Local Similarity 49.1%; Pred. No. 0.00012;
Matches 213; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 1254 cattggtcagctcaatttaattgaggagggttgtaacgcgtcgtgacgaagtag 1313
DB 57 CTTTAATATTCTATATGAAGAAATGGAATGATAATATATGCAATTTATCCGATG 116
QY 1314 aaagtaacaattcaacggtgtgtgctgaagttaactgaacgtcaagctaaatggtcaaa 1373
DB 117 CAAATAATAATGAATAAANAATAATCAAAATAAATAAATAATATGAAATAATAATAATA 176
QY 1374 ttggcgtagatggaaaaattattagctgaacagtaatacaacagcgtgctgcgaacaatc 1433
DB 177 ATACAAACAATAATAATAACAATAATAATAATAATAATAATAATAATAATAATAATA 236
QY 1434 aagaatcatctgctactaacaactactgaaatgctgtgttaatgaaagcaaaaaacta 1493
DB 237 ACAATAATAATAGCAACAATAATAATAATAGCAACAATAATAATAATAATAATAATA 296
QY 1494 acaactgaaaatgctgtgttgaatgaaacaaaaataactaacaatactgaaaatgctg 1553
DB 297 ATAGTACACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 356
QY 1554 ttgttaatgaaacaaaaataactaacaacacacacacacacacacacacacacacacac 1613
DB 357 ATAGTACACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 416
QY 1614 ctaaatgaacacaccatcagcgtcactcaagcaaacactgaagcgtgaaagcctcctcaag 1673
DB 417 ATAATAATACATAATAGTAAACAATAATAATAATAATAATAATAATAATAATAATAATA 476
QY 1674 ctgctaaggctgtt 1687
DB 477 ATGATAATACTATT 490

RESULT 11
BH391959 773 bp DNA linear GSS 11-DEC-2001
LOCUS AG-ND-180F9.TF ND-TAM Anopheles gambiae genomic clone AG-ND-180F9,
DEFINITION DNA sequence.

ACCESSION BH391959
VERSION BH391959.1 GI:17338087
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Pterygota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
; Anopheles.
1 (bases 1 to 773)
Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSs: AG-ND-180F9.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F. H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1..773
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-180F9"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 309 a 125 c 99 g 240 t
ORIGIN

Query Match 1.7%; Score 81; DB 12; Length 773;
Best Local Similarity 57.2%; Pred. No. 0.00012;
Matches 147; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 1363 taatggtcaaatggcgtagatggaaaaattattagtgtaacagtaatacaaccagtgg 1422
DB 377 TAATAATAATAATAAACAATAATAATAATAATAATAATAATAATAATAATAATAATAA 436
QY 1423 ctgcacaaatcaagaatcatctgctactaacaactactgaaatgctgtgttaatgaaag 1482
DB 437 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 496
QY 1483 caaaataactaacaatactgaaatgctgtgttaatgaaacaaaaataactaacaatac 1542
DB 497 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 556
QY 1543 tgaataatgctgtgttaatgaaacaaaaataactaacaacacacacacacacacacacac 1602
DB 557 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 616
QY 1603 attaaagttaactaata 1619
DB 617 TCTTACTAATAATAAAAA 633

RESULT 12
AQ656286
LOCUS Sheared DNA-27K23.TF Sheared DNA Trypanosoma brucei genomic clone
DEFINITION Sheared DNA-27K23, DNA sequence.

RESULT 14
 A0656218/c
 LOCUS
 DEFINITION
 639 bp DNA linear GSS 23-JUN-1999
 Sheared DNA-27123.TF Sheared DNA Trypanosoma brucei genomic clone
 A0656218
 VERSION
 A0656218.1 GI:5163986
 GSS.
 ORGANISM
 Trypanosoma brucei.
 Trypanosoma brucei
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.
 1 (bases 1 to 639)
 El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
 Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
 Fraser,C. and Adams,M.
 Determination of clone end sequences from Trypanosoma brucei GUTat
 10.1 sheared DNA library
 Unpublished (1999)
 Other_GSSs: Sheared DNA-27123.TF
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through ATCC. Sheared DNA end sequences search page:
 http://www.tigr.org/tldb/mdb/tbdb/.
 Seq primer: M13-Forward
 Class: shotgun.

FEATURES
 source
 Location/Qualifiers
 1..639
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone_lib="Sheared DNA-27123"
 /note="Vector: pUC18; Site_1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Vaudin and B. Borell, Oxford University
 Press, 1999)."
 197 a 94 c 41 g 307 t

Query Match
 Best Local Similarity 1.7%; Score 79.4; DB 12; Length 639;
 Matches 140; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
 QY 1383 atggaaaattattagtctaacagtaacacaccagtgctgcgacaaatacaagaatcat 1442
 DB 547 AGGGCAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 488
 QY 1443 ctgtactacaatactgaaatgctgtgttaataagaagcaaaataactacaactg 1502
 DB 487 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 428
 QY 1503 aaaaatgctgtgttaatagaacaaaataactacaatactgaaatgctgtgttaatg 1562
 DB 427 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 368

QY 1563 aaacaaataactaacacacagataatagtcataatgtaagttactaataatg 1622
 DB 367 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATG 308
 QY 1623 a 1623
 DB 307 A 307

RESULT 15
 A0658275/c
 LOCUS
 DEFINITION
 560 bp DNA linear GSS 23-JUN-1999
 Sheared DNA-3M5.TR Sheared DNA Trypanosoma brucei genomic clone
 A0658275
 VERSION
 A0658275.1 GI:5166043
 GSS.
 ORGANISM
 Trypanosoma brucei.
 Trypanosoma brucei
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.
 1 (bases 1 to 560)
 El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
 Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
 Fraser,C. and Adams,M.
 Determination of clone end sequences from Trypanosoma brucei GUTat
 10.1 sheared DNA library
 Unpublished (1999)
 Other_GSSs: Sheared DNA-3M5.TF
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through ATCC. Sheared DNA end sequences search page:
 http://www.tigr.org/tldb/mdb/tbdb/.
 Seq primer: M13-Reverse
 Class: shotgun.

FEATURES
 source
 Location/Qualifiers
 1..560
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone_lib="Sheared DNA-3M5"
 /note="Vector: pUC18; Site_1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Vaudin and B. Borell, Oxford University
 Press, 1999)."
 163 a 78 c 28 g 291 t

Query Match
 Best Local Similarity 1.7%; Score 79.2; DB 12; Length 560;
 Matches 138; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
 QY 1388 aaaaattattagtctaacagtaacacaccagtgctgcgacaaatacaagaatcatcgt 1447
 DB 556 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 497
 QY 1448 actaacaatactgaaatgctgtgttaataagaagcaaaataactacaatactgaaat 1507

Search completed: September 22, 2002, 01:51:41
Job time: 8381 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 05:22:06 ; Search time 180.95 Seconds
(without alignments)
6290.498 Million cell updates/sec

Title: US-09-995-587A-10
Perfect score: 4634
Sequence: 1 gtaacaaagacaaatttt.....gttaactgctgcacaaagctt 4634

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	75	1.6	3305	3	US-09-068-043-1
2	69.4	1.5	10317	3	US-09-058-746-1
3	69.4	1.5	10317	4	US-09-438-142-1
4	64.4	1.4	740	4	US-09-451-117-1
5	61.2	1.3	1365	2	US-08-870-827-4
6	61.2	1.3	2408	2	US-08-870-827-5
7	59.8	1.3	4507	2	US-08-568-459A-3
8	59.8	1.3	4507	2	US-08-487-826B-3
9	57.2	1.2	3504	2	US-08-760-797A-2
10	57.2	1.2	3504	4	US-08-932-929B-2
11	56.6	1.2	8920	2	US-08-446-855A-1
12	56.6	1.2	8920	4	US-09-150-741-1
13	56.2	1.2	1631	3	US-09-118-319-1
14	55.4	1.2	3504	2	US-08-760-797A-4
15	55.4	1.2	3504	4	US-08-932-929B-4
16	54.4	1.2	9636	1	US-08-323-170B-1
17	54.4	1.2	9636	4	US-08-954-441-1
18	53.4	1.2	2237	4	US-08-914-999-7
19	52.4	1.1	1110	1	US-08-257-073-11
20	51.2	1.1	7218	1	US-08-232-463-14
21	51.2	1.1	19124	2	US-08-487-826B-13
22	50.4	1.1	1243	3	US-09-009-620-1
23	50.4	1.1	2223	1	US-08-257-073-4
24	50.4	1.1	4896	4	US-09-210-361-3
25	50.2	1.1	5116	1	US-08-038-682-1
26	50.2	1.1	5116	1	US-08-302-832-1
27	50.2	1.1	5116	2	US-08-530-198-1

28	50.2	1.1	5116	2	US-08-469-880-1	Sequence 1, Appl
29	50.2	1.1	5116	2	US-08-728-470-1	Sequence 1, Appl
30	50.2	1.1	5116	2	US-08-617-697-1	Sequence 1, Appl
31	50.2	1.1	5116	4	US-08-719-641-1	Sequence 1, Appl
32	50.2	1.1	9171	1	US-08-038-682-5	Sequence 5, Appl
33	50.2	1.1	9171	1	US-08-302-832-5	Sequence 5, Appl
34	50.2	1.1	9171	2	US-08-530-198-5	Sequence 5, Appl
35	50.2	1.1	9171	2	US-08-469-880-5	Sequence 5, Appl
36	50.2	1.1	9171	2	US-08-728-470-5	Sequence 5, Appl
37	50.2	1.1	9171	2	US-08-617-697-5	Sequence 5, Appl
38	50.2	1.1	9171	4	US-08-719-641-5	Sequence 5, Appl
39	49.4	1.1	1956	4	US-08-559-896B-1	Sequence 1, Appl
40	48.2	1.0	665	2	US-08-883-795A-36	Sequence 36, Appl
41	48.2	1.0	5852	1	US-07-867-106-2	Sequence 2, Appl
42	47.8	1.0	6744	1	US-08-119-125A-2	Sequence 2, Appl
43	47	1.0	6124	4	US-08-213-419B-3	Sequence 3, Appl
44	46.8	1.0	4248	3	US-08-678-614-1	Sequence 1, Appl
45	46.6	1.0	1431	4	US-09-316-083-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-068-043-1
; Sequence 1, Application US/09068043
; Patent No. 6048694
; GENERAL INFORMATION:
; APPLICANT: MICHAEL GENE BRAMUCCI
; APPLICANT: VASANTHA NAGARAJAN
; TITLE OF INVENTION: POSITIVE SELECTION
; TITLE OF INVENTION: VECTOR FOR BACILLUS SP.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND
; ADDRESSEE: COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT OFFICE 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/068,043
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,201
; FILING DATE: NOVEMBER 3, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FELTHAM, S. NEIL
; REGISTRATION NUMBER: 36,506
; REFERENCE/DOCKET NUMBER: CR-9807
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-6460
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; US-09-068-043-1

Query Match 1.6%; Score 75; DB 3; Length 3305;
Best Local Similarity 50.7%; Pred. No. 9.8e-09;

```
Matches 264; Conservative 0; Mismatches 245; Indels 12; Gaps 3;
QY 2132 acaaatgttccaagaataggtcaggttcaggtatgtataatgatgtgtacaaattcaa 2191
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2273 AAAGATCAGACGGAAGAATGGTCCGTTCTGCGAACCTTTTACATCTGACGGAAAAATCCGT 2332
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2192 ctattttccacctcaaatgatacgtctgattacaagtgtgaatgatacgccttgcacc 2251
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2333 TTATCTACAC-----TGACTATTCCGGTAAACATTTACGGCAACAAAGCCTGACAAACA 2386
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2252 gcaacattaaaccttaagtgttgatgatacgggtgtttcaatcaagagtggttgataattat 2311
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2387 GCGCAGGTAATGTGTCAAAATCTGATGACACACTCAAAATCAACGGAGTGGAGAGATCAC 2446
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2312 caagttttgttgaagtgatgatttcactaccacaaacttatgaacaattcgcaaacggc 2371
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2447 AAGCAGATTTTTCAGCGAGACGGCAAAACATATCAGAACGTTTCAGCAGTTTATCGATGAA 2506
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2372 aaagatcgtgaaatgatgattactgttactgtgacccacacggtgttcaattagaaaat 2431
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2507 GGAATATTATACATCCGCGACACACCATACGCTGAGAGACCTCTACTACGTTGAAGACAAA 2566
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2432 ggtgatcgttatcttatttcgaagcttaactctgggacagaa---gattaccacaaagtgc 2488
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2567 GGCATAAATACCTTGTATTTCGAAGCAACACACGGGAACAGAAACGGATACCAAGGCGAA 2626
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2489 gacaaatttataattgggctaactatggtggcgatgagtccttcaattataagagttcc 2548
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2627 GAATCTTTATTAAACAAAGCGTACTACGGGGCGGACGCAACTTCTCCCTAAAGAAAGC 2686
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2549 ttcaagcttttgaataataaagaagatcgtgaattggctggtttagctaatgtgacctt 2608
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2687 CAGAGCTTCAGCAGACGCGCTAAACACGGA---TGCTGAGTACGCAACGGCGCCCTC 2743
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2609 ggtattttaaagtcactaacaatcaataaagtaagccaaaggt 2649
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2744 GGTATCATAGAGTTAAATAATGATTACACATTGAAAAAAGT 2784
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-09-058-746-1/c
; Sequence 1, Application US/09058746
; Patent No. 6022716
; GENERAL INFORMATION:
; APPLICANT: Ilya Chumakov
; APPLICANT: Hiroaki Tanaka
; TITLE OF INVENTION: High Throughput DNA sequencing vector
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,746
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.015A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10317 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: CIRCULAR
; MOLECULE TYPE: synthetic DNA
; ORGANISM: Cloning vector pGendel
; FEATURE:
; NAME/KEY: pGendel
; LOCATION: 1..10317
; FEATURE:
; NAME/KEY: Homology with X06404 compl (411..1668)
; LOCATION: 9..1268
; IDENTIFICATION METHOD: blastn against X06404
; FEATURE:
; NAME/KEY: Kanamycin resistance gene CDS
; LOCATION: 142..957
; IDENTIFICATION METHOD: By homology to X06404
; FEATURE:
; NAME/KEY: Tn1000's right end
; LOCATION: complement 1332..1371
; IDENTIFICATION METHOD: blastn against X60200)
; FEATURE:
; NAME/KEY: Homology with U46017 (1-472)
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: blastn against U46017
; FEATURE:
; NAME/KEY: single stranded DNA replication origin
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: By homology to U46017
; OTHER INFORMATION: mutation T -> C 1658
; FEATURE:
; NAME/KEY: Homology with U51113 (2382..6997)
; LOCATION: 1896..6544
; IDENTIFICATION METHOD: blastn against U51113
; FEATURE:
; NAME/KEY: OriS
; LOCATION: 1972..2188
; IDENTIFICATION METHOD: By homology to U51113
; FEATURE:
; NAME/KEY: repELR
; LOCATION: 2897..2918
; OTHER INFORMATION: Described in seqID 16
; FEATURE:
; NAME/KEY: RepE
; LOCATION: 2903..3034
; IDENTIFICATION METHOD: By homology to U51113
; FEATURE:
; NAME/KEY: T3
; LOCATION: 3043..3059
; OTHER INFORMATION: Described in seqID 17
; FEATURE:
; NAME/KEY: LRT3RA
; LOCATION: complement 3045..3069
; OTHER INFORMATION: Described in seqID 15
; FEATURE:
; NAME/KEY: IncC
; LOCATION: 3070..3320
; IDENTIFICATION METHOD: By homology to U51113
; OTHER INFORMATION: Insertion 33 bases 3038..3071
; FEATURE:
; NAME/KEY: Para
; LOCATION: 3655..4821
; IDENTIFICATION METHOD: By homology to U51113
; OTHER INFORMATION: mutation G -> A 3878
; FEATURE:
; NAME/KEY: ParB
; LOCATION: 4821..5792
; IDENTIFICATION METHOD: By homology to U51113
; FEATURE:
; NAME/KEY: ParC
; LOCATION: 5865..6382
; IDENTIFICATION METHOD: By homology to U51113
```

```

FEATURE:
NAME/KEY: Homology with J01688 (complement 175..819)
LOCATION: 6574..7218
IDENTIFICATION METHOD: blastn against J01688
OTHER INFORMATION: mutation A -> G 7096
FEATURE:
NAME/KEY: CDS streptomycin sensitivity gene
LOCATION: complement 6716..7090
IDENTIFICATION METHOD: By homology to J01688
OTHER INFORMATION: mutation A -> G 6728
OTHER INFORMATION: mutation G -> C 6821
OTHER INFORMATION: mutation C -> T 6866
OTHER INFORMATION: mutation T -> C 7013
OTHER INFORMATION: mutation T -> A 7058
FEATURE:
NAME/KEY: rpsLR
LOCATION: 7155..7174
OTHER INFORMATION: Described in seqID 12
FEATURE:
NAME/KEY: SP6
LOCATION: 7230..7248
OTHER INFORMATION: Described in seqID 13
FEATURE:
NAME/KEY: Tn1000's left end
LOCATION: 7252..7291
IDENTIFICATION METHOD: blast (X60200)
FEATURE:
NAME/KEY: Homology with X02730 (complement 37..1959)
LOCATION: 7305..9227
IDENTIFICATION METHOD: blastn against X02730
FEATURE:
NAME/KEY: CDS levansucrase gene
LOCATION: complement 7379..8800
IDENTIFICATION METHOD: By homology to X02730
OTHER INFORMATION: mutation T -> C 7466
OTHER INFORMATION: mutation A -> G 7739
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347
OTHER INFORMATION: mutation T -> C 8600
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772
FEATURE:
NAME/KEY: SLR3
LOCATION: 8711..8731
OTHER INFORMATION: Described in seqID 14
FEATURE:
NAME/KEY: Homology with J01636 (complement 1158..1465)
LOCATION: 9298..9623
IDENTIFICATION METHOD: blastn against J01636
FEATURE:
NAME/KEY: CDS alpha peptide beta-galactosidase
LOCATION: complement 9276..9497
IDENTIFICATION METHOD: By homology to J01636
FEATURE:
NAME/KEY: primer HE1
LOCATION: complement 9465..9479
FEATURE:
NAME/KEY: primer HE2
LOCATION: 9461..9475
FEATURE:
NAME/KEY: primer LacLR2Avr
LOCATION: complement 9603..9630
FEATURE:
NAME/KEY: primer LacE2Mlu
LOCATION: 9289..9314
FEATURE:
NAME/KEY: Homology with M7789 (1889..2576)
LOCATION: 9629..10315
IDENTIFICATION METHOD: blastn against M7789
FEATURE:
NAME/KEY: high copy-number double-stranded DNA replication origin
LOCATION: complement 9629..10315
IDENTIFICATION METHOD: By homology to M7789
OTHER INFORMATION: mutation C -> T 9803
OTHER INFORMATION: site ScaI 10029 - 10034

```

```

OTHER INFORMATION: site PmlI 10038 - 10043
OTHER INFORMATION: CLONING SITES 10031 - 10041
FEATURE:
NAME/KEY: oriLRd
LOCATION: 9856..9881
OTHER INFORMATION: Described in seqID 8
FEATURE:
NAME/KEY: OS1
LOCATION: 10009..10026
OTHER INFORMATION: Described in seqID 10
FEATURE:
NAME/KEY: OR1
LOCATION: complement 10046..10062
OTHER INFORMATION: Described in seqID 11
FEATURE:
NAME/KEY: oriLRr
LOCATION: complement 10182..10202
OTHER INFORMATION: Described in seqID 9
US-09-058-746-1

Query Match      1.5%; Score 69.4; DB 3; Length 10317;
Best Local Similarity 47.4%; Pred. No. 3.4e-07;
Matches 381; Conservative 0; Mismatches 392; Indels 30; Gaps 5;

QY 2144 caagaatggtcaggttcagctattgttaaatgatggtgtacaaattcaactatttttccacc 2203
DB 8320 CAAGAATGTCAGGTCAGCCACATTTACATCTGCGGAAAAATCGGTTATTCTACAC- 8360
QY 2204 tcaaatgatacgtctgtattacaaagtgaatgatacgaagcttctaccgcaacattaaac 2263
DB 8261 -----TGATTTCTCCGGTAAACATTTACGGCAAAACAACACTGACAACTGCACAAGTTAAC 8207
QY 2264 cttaattgtgatgaacaggtgtttcaatcaagaggtgtgataattatcaagtttggttt 2323
DB 8206 GTATCAGCATCAGACAGCTCTTTTGAACATCACGGGTGAGAGGATTATAAATCAATCTTTT 8147
QY 2324 gaaggtgatgttttcactaccacaaacttatgaacaatttcgcaaacggcgaagatcgtgaa 2383
DB 8146 GACGGTGACGGAAAAACAGTATCAAAATGTACAGCAGTTTCATCGATGAGCAACTACAGC 8087
QY 2384 aatgatgattactgtcttaactgacccacacgctgttccaattagaataatggtgactgtat 2443
DB 8086 TCAGCGCACAAACCATACACGCTGAGAGATCTCTACTAGTAGAAGATAAAGGCCACAAATAC 8027
QY 2444 ctgtatttcgaagctaaatactgggacagaagat---taccaaagtacacacaaatttat 2500
DB 8026 TTAGTATTGTAAGCAAAACACTGGAACCTGAAGATGGCTACCAAGCGCAAGAAATCTTTATT 7967
QY 2501 aattgggttaactatggtggcgtgatgatgccttcaatattaaagagttccttcaagctttg 2560
DB 7966 AACAAAGCATACTATGCGGAAAAAGCACATCATCTTCCGTCAGAAAGTCAAAAACCTCTG 7907
QY 2561 aataataagaagatcgtgaattgctgtttagcttaagtgtgacttggtgacttcttaag 2620
DB 7906 CAAAGCGATAAAAAACCGCA---CGGCTGAGTTAGCAAAACGGCGCTCTCGGTATGATTGAG 7850
QY 2621 ctcaactaacaatcaaaagtaagcgaaggttgaagaagtatactcaccattggttatctact 2680
DB 7849 CTTAAACGATGATTACA-----CACTGAAAAAAGTATGAACCCGCTGATGTCATCTAAC 7796
QY 2681 ttgatggcttgcgtgaggtannnnnnaagcttgggtgaataagtattatctcttccgta 2740
DB 7795 ACAGTAACAGATGAATTAACACGCGCAACGCTCTTTAAAAATGAACGGCAAAATGGTATCTG 7736
QY 2741 actcgttaagtcgtggttccgatcgtgaattaacgcgttaaggaataacacacatcgttgg 2800
DB 7735 TTCACCTGA-----CTCCCGCGGATCAAAAATGACGATTGACGGCATTACGTCT 7688
QY 2801 gataacggttgcgtatgattggttacgttttccgatagcttaagtgggaagacacagccatta 2860
DB 7687 AACGATATTACATGCTGGTTATGTTTCTTAATCTTTTAACCTGGCCCATACAAAGCCGCTG 7628

```

```

Oy 2861 aataactcaggtgtcgattactgacatcagctacacgcaaacgtgacgtactgactactat 2920
    || || || || || || || || || || || || || || || || || || || || || ||
Db 7627 AACAAACTGGCCCTTGTGTTAAAAATGGATCTTGATCCTCAACGATGTAACCTTTACTTAC 7568
    || || || || || || || || || || || || || || || || || || || || || ||
Oy 2921 tctactatgacgtacctgtgacg 2943
    || || || || || || || || || || || || || || || || || || || || || ||
Db 7567 TCACACTTCGCTGTACCTCAAGC 7545

RESULT 3
US-09-438-142-1/c
; Sequence 1, Application US/09438142
; Patent No. 6258571
; GENERAL INFORMATION:
; APPLICANT: Ilya Chumakov
; APPLICANT: Hiroaki Tanaka
; TITLE OF INVENTION: High Throughput DNA sequencing vector
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/438,142
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.015C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10317 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: synthetic DNA
; ORIGINAL SOURCE:
; ORGANISM: Cloning vector pGendel
; FEATURE:
; NAME/KEY: pGendel
; LOCATION: 1..10317
; FEATURE:
; NAME/KEY: Homology with X06404 compl (411..1668)
; LOCATION: 9..1266
; IDENTIFICATION METHOD: blastn against X06404
; FEATURE:
; NAME/KEY: Kanamycin resistance gene CDS
; LOCATION: 142..957
; IDENTIFICATION METHOD: By homology to X06404
; FEATURE:
; NAME/KEY: Tn1000's right end
; LOCATION: complement 1332..1371
; IDENTIFICATION METHOD: blastn against X02000)
; FEATURE:
; NAME/KEY: Homology with U46017 (1-472)
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: blastn against U46017
; FEATURE:
; NAME/KEY: single stranded DNA replication origin
; LOCATION: 1423..1894
; IDENTIFICATION METHOD: By homology to U46017
; OTHER INFORMATION: mutation T -> C 1658
; FEATURE:
; NAME/KEY: Homology with U51113 (2382..6997)
; LOCATION: 1896..6544
; IDENTIFICATION METHOD: blastn against U51113
; FEATURE:
; NAME/KEY: OriS
; LOCATION: 1972..2188
; IDENTIFICATION METHOD: By homology to U51113
; FEATURE:
; NAME/KEY: repELR
; LOCATION: 2897..2918
; OTHER INFORMATION: Described in seqID 16
; FEATURE:
; NAME/KEY: RepE
; LOCATION: 2903..3034
; IDENTIFICATION METHOD: By homology to U51113
; FEATURE:
; NAME/KEY: T3
; LOCATION: 3043..3059
; OTHER INFORMATION: Described in seqID 17
; FEATURE:
; NAME/KEY: LRT3RA
; LOCATION: complement 3045..3069
; OTHER INFORMATION: Described in seqID 15
; FEATURE:
; NAME/KEY: IncC
; LOCATION: 3070..3320
; IDENTIFICATION METHOD: By homology to U51113
; OTHER INFORMATION: insertion 33 bases 3038..3071
; FEATURE:
; NAME/KEY: Para
; LOCATION: 3655..4821
; IDENTIFICATION METHOD: By homology to U51113
; OTHER INFORMATION: mutation G -> A 3878
; FEATURE:
; NAME/KEY: ParB
; LOCATION: 4821..5792
; IDENTIFICATION METHOD: By homology to U51113
; FEATURE:
; NAME/KEY: ParC
; LOCATION: 5865..6382
; IDENTIFICATION METHOD: By homology to U51113
; FEATURE:
; NAME/KEY: Homology with J01688 (complement 175..819)
; LOCATION: 6574..7218
; IDENTIFICATION METHOD: blastn against J01688
; OTHER INFORMATION: mutation A -> G 7096
; FEATURE:
; NAME/KEY: CDS streptomycin sensitivity gene
; LOCATION: complement 6716..7090
; IDENTIFICATION METHOD: By homology to J01688
; OTHER INFORMATION: mutation A -> G 6728
; OTHER INFORMATION: mutation G -> C 6821
; OTHER INFORMATION: mutation C -> T 6866
; OTHER INFORMATION: mutation T -> C 7013
; OTHER INFORMATION: mutation T -> A 7058
; FEATURE:
; NAME/KEY: fplSLR
; LOCATION: 7155..7174
; OTHER INFORMATION: Described in seqID 12
; FEATURE:
; NAME/KEY: SP6
; LOCATION: 7230..7248
; OTHER INFORMATION: Described in seqID 13
; FEATURE:
; NAME/KEY: Tn1000's left end
; LOCATION: 7252..7291
; IDENTIFICATION METHOD: blast (X60200)
; FEATURE:
; NAME/KEY: Homology with X02730 (complement 37..1959)
; LOCATION: 7305..9227
; IDENTIFICATION METHOD: blastn against X02730
; FEATURE:

```



```

Query Match      1.4%; Score 64.4; DB 4; Length 740;
Best Local Similarity 50.3%; Pred. No. 1.9e-06;
Matches 158; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 1287 ttgtaaccgtctcatgcgtacatcaagtagaaagttaaacaattacaacgggtgttctgctaagaagtta 1346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 ttacatctagtggtgatctcctagctgtttactacaataactcttaagtgtgaataataatagta 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1347 tactgaacgtcaagcgtaagtgtcacaattggcgtagatggaaaaattattattagtgcttaaca 1406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 atagtaatatggtatcaggaaaattttataccagttggtacctgttcttctactagtattg 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1407 gtaatacacaccagtgctgcagacaatacaagaatcatctgctactaacaaataactgaaaatg 1466
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 gttaatgtaatgggtgtgctttactgtcttatccttcataataataacaatagcataata 604
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1467 ctgtgtttaatgaagcaaaaatactacaatactgaaaatgctgttgtttaatgaaaaca 1526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 ttaataataataataataatagtaataccactcttactactgtgtgtgtactaatgcta 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1527 aaaataactacaataactgaaaatgctgtgtttaatgaaaacaaaaataactaacacacag 1586
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 665 attactactactaactactaatactactactactactactactactactactaataataataata 724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1587 aaaaagcataatagt 1600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 725 ataataataataat 738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-08-870-827-4
; Sequence 4, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1365 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-870-827-4

```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sp.
; INDIVIDUAL ISOLATE: V230 (FERM BP-5054)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..360
; IDENTIFICATION METHOD: E
; NAME/KEY: signal peptide
; LOCATION: 361..456
; IDENTIFICATION METHOD: S
; NAME/KEY: mat peptide
; LOCATION: 457..1821
; IDENTIFICATION METHOD: S
; NAME/KEY: 3'UTR
; LOCATION: 1822..2408
; IDENTIFICATION METHOD: E
; US-08-870-827-5

Query Match          1.3%; Score 61.2; DB 2; Length 2408;
Best Local Similarity 50.0%; *Pred. No. 1.8e-05;
Matches 181; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 2299 tgttgataattcaagttttgttggaagtgatgatttcactaccacaaacttatgaaca 2358
Db 1047 TGATCATAAATCTGCTTTTGATGGCGGAGACGGTACAGTTTATCAAAATATTCAGCAATT 1106

QY 2359 attcgcaaacgcaagatcgigaaatgatgattactctgactgacccacacattgt 2418
Db 1107 TATCGATGAAGCAAGTGCATTCAGGTGATAACCATCTTTAAGAGACCCCTCATCTATGT 1166

QY 2419 tcaattagaaagtgatgctgttatttcgaagctaatactgggacagaaattat 2478
Db 1167 TGA---AGATAAGGGCCATAAATATCTTCTTTGAAGCGAATACTGGAACAACACATGG 1223

QY 2479 ccaagtgacgaccaaattataattgggctaactatggcgcatgacgcttcacat 2538
Db 1224 TTATCAAGCGGATGACTCTTTCAATAAATAAGTCTTATGCGGGAAGTGACGTCTCTT 1283

QY 2539 taaggtctcttcgaagcttttgaataataagaagatcgatgctgaattggctgttagctaa 2598
Db 1284 CCAGAAATGAAAAATAAACTGCTTCAAAGTCTTAAARACAAATGCTCTTTAGCGAA 1343

QY 2599 tgggtcaacttgatcttaaaagctcactaacaatacaagtaagccaaaggttgaagaagt 2658
Db 1344 TGGTGCAATTAGGCATGTTTGAATTGGCCGATGACTATACAGTGAAGAAAGTGTATGAAC 1403

QY 2659 at 2660
Db 1404 AT 1405

RESULT 7
US-08-568-459A-3
; Sequence 3, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Welles, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
```

```
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4507 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-3
```

```
Query Match          1.3%; Score 59.8; DB 2; Length 4507;
Best Local Similarity 46.0%; Pred. No. 5e-05;
Matches 202; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 1308 aagtagaagaaagtaacattcaaacgggtgttgctgaagttaactactgaacgccaagctaatg 1367
Db 2670 AAGATCAAAATAGTGTATGAATCGGAAGAAACTGTACTAATATATATCAAAAGTCCAT 2729

QY 1368 gtcaaatggcgtagatgaaataattattagtgctaacagtaatacaacccagtggtcga 1427
Db 2730 CTATAAATAATGGAGATGATTGAGGAGTGAAGTCAACAGTGAATCTAGTAGTT 2789

QY 1428 caaatcaagaatcatctgctactaacaataactgaaagtctgtgttgaatgaagcaaaa 1487
Db 2790 CAAATACTGGATTGCTATTGATGATGATAGAAATGGTGATACATTTGTTGACACACAAG 2849

QY 1488 atactaacaatactgaaagtgtgtgttgaatgaacacacacacacacacacacacacac 1547
Db 2850 ATACAGCAAAATACTGAAGATGTTATTAGAAAAGAAAATGCTGCAAGGATGAAGATGAAA 2909

QY 1548 atgctgtgttgaatgaacacacacacacacacacacacacacacacacacacacacac 1607
Db 2910 AAGGCGCAGATGAAGAAGACATAGTACTTCTGAAAGCTTAAGTTCACTGAAAGAAAAA 2969

QY 1608 agttaactaataatgaacacacacacacacacacacacacacacacacacacacacacac 1667
Db 2970 TGTAACTGATAAATGAAGGAGGAAATAGTTTAAATCATGAAGAGGTGAAGAAGACATACTA 3029

QY 1668 ctcaagctgctaaagctgttcaaaatgccaagattgatgcccggttttaacagatgatc 1727
Db 3030 GTAATCTGATAAATGTTCAACAGTCTGGAGGAATTTGTTAATATGATGTTGAGAAAGAAC 3089

QY 1728 aaattaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1746
Db 3090 TAAAGATCTTTAGAAATA 3108

RESULT 8
us-08-487-826B-3
; Sequence 3, Application US/08487826B
; Patent No. 5993827
```


Db 686 catcttaaggctgctggttcttaaaactgttgta 718

RESULT 14

US-08-760-797A-4
; Sequence 4, Application US/08760797A
; Patent No. 5928902
; GENERAL INFORMATION:
; APPLICANT: De Wilde, Michel
; APPLICANT: Cohen, Joseph
; TITLE OF INVENTION: Hybrid Protein Between CS
; TITLE OF INVENTION: Hybrid Protein Between CS
; TITLE OF INVENTION: from Plasmodium and HBSAG
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/760.797A
; FILING DATE: 04-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/442,612
; FILING DATE: 17-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: B45015-1C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3504 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-760-797A-4

Query Match 1.2%; Score 55.4; DB 2; Length 3504;
Best Local Similarity 47.0%; Pred. No. 0.00052;
Matches 213; Conservative 0; Mismatches 231; Indels 9; Gaps 1;
QY 1443 ctgctactaacaatactgaaatgctgtgttaataagcaaaataactaactaactg 1502
Db 1069 CCATCTTAATCAATCCAAATCAAAACCAACCAACCAACCAACCAATCAATCAATG 1128
QY 1503 aaaaatgctgtgttaataagcaaaataactaactaactgaaatgctgtgttaatg 1562
Db 1129 CAACCCCAATGCAATCTTAATGCAATCTTAATGCAATCTTAATGCAATCTTAATG 1188
QY 1563 aaaaataactaactaactaactaactaactaactaactaactaactaactaactg 1622
Db 1189 CAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAATG 1248
QY 1623 aaaaatgctgtgttaataagcaaaataactaactaactgaaatgctgtgttaatg 1673
Db 1249 CAACCCCAATGCAATCTTAATGCAATCTTAATGCAATCTTAATGCAATCTTAATG 1308
QY 1674 ctgtaaggctgttcaaaatgcaaaatgcaaaatgcaaaatgcaaaatgcaaaatg 1733
Db 1309 GTATGACAAAGGTGACAAATATGCAATATGCAATATGCAATATGCAATATGCAATATG 1368

QY 1734 atgaattaaataagattacttctcttaagctgctgaaaggggtgcaaaatgacattca 1793
Db 1369 ATGCCAACAGTCTGCTATAAAATAATAATACGAGAACCAAGTATAGACATATAAAG 1428
QY 1794 aggaacttagagggtgtgttaagctgtattgttaagcaagatccacaatatgtattcctt 1853
Db 1429 ATATTTTAAACAAATAACAAAATCTCTTTCAACTGAATGGTCCCATGCTAGTGAACCT 1488
QY 1854 attctaagcttaaggaatacaagaataatgcctg 1886
Db 1489 GTGGAATGGTATTCAAGTTAGATAAAGCCTG 1521

RESULT 15

US-08-932-929B-4
; Sequence 4, Application US/08932929B
; Patent No. 6169171
; GENERAL INFORMATION:
; APPLICANT: De Wilde, Michel
; APPLICANT: Cohen, Joseph
; TITLE OF INVENTION: Hybrid Protein Between CS
; TITLE OF INVENTION: from Plasmodium and HBSAG
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,929B
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/760,797
; FILING DATE: 04-DEC-1996
; APPLICATION NUMBER: 08/442,612
; FILING DATE: 17-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: B45015-1FWC2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3504 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-932-929B-4

Query Match 1.2%; Score 55.4; DB 4; Length 3504;
Best Local Similarity 47.0%; Pred. No. 0.00052;
Matches 213; Conservative 0; Mismatches 231; Indels 9; Gaps 1;
QY 1443 ctgctactaacaatactgaaatgctgtgttaataagcaaaataactaactaactg 1502
Db 1069 CCATCTTAATCAATCCAAATCAAAACCAACCAACCAACCAACCAATCAATCAATG 1128
QY 1503 aaaaatgctgtgttaataagcaaaataactaactaactgaaatgctgtgttaatg 1562
Db 1129 CAACCCCAATGCAATCTTAATGCAATCTTAATGCAATCTTAATGCAATCTTAATG 1188

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2002, 09:38:17 ; Search time 1031.3 Seconds
(without alignments)
7714.711 Million cell updates/sec

Title: US-09-995-587A-10
Perfect score: 4634
Sequence: 1 gttacaagacaaaatttt.....gttactgctgcaaaagctt 4634

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues 3472872

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.6	1.9	2079	23	NAAS85786
c	88.6	1.9	2079	23	NAAS85786
3	88.6	1.9	2079	23	NAAS87342
c	88.6	1.9	2079	23	NAAS87342
5	80.8	1.7	5940	21	NAAS88703
6	78.2	1.7	3947	23	NAAS86050
c	77.2	1.7	6642	23	NAAS81427
c	76.6	1.7	979	23	NAAS86024
c	76	1.6	5340	22	NAAF26289

10	75.6	1.6	3990	23	AAS89967	DNA encoding novel	
11	75.6	1.6	3990	23	AAS92589	DNA encoding novel	
12	75.6	1.6	6033	23	AAS85839	DNA encoding novel	
13	75.4	1.6	1866	23	AAS86014	DNA encoding novel	
14	75.4	1.6	1866	23	AAS87349	DNA encoding novel	
15	75	1.6	1032	23	AAS86020	DNA encoding novel	
16	75	1.6	1099	23	AAS74868	DNA encoding novel	
17	75	1.6	1188	23	AAS68479	DNA encoding novel	
18	75	1.6	1188	23	AAS74483	DNA encoding novel	
19	75	1.6	1188	23	AAS86028	DNA encoding novel	
20	75	1.6	1188	23	AAS87344	DNA encoding novel	
c	21	75	1.6	1212	23	AAS86040	DNA encoding novel
	22	75	1.6	1245	23	AAS86037	DNA encoding novel
	23	75	1.6	1281	23	AAS68475	DNA encoding novel
	24	75	1.6	1281	23	AAS87378	DNA encoding novel
25	75	1.6	1284	23	AAS65782	DNA encoding novel	
26	75	1.6	1284	23	AAS74491	DNA encoding novel	
27	75	1.6	1453	23	AAS69429	DNA encoding novel	
28	75	1.6	1459	23	AAS85820	DNA encoding novel	
c	29	75	1.6	1512	23	AAS86017	DNA encoding novel
	30	75	1.6	1512	23	AAS88769	DNA encoding novel
c	31	75	1.6	1535	23	AAS86038	DNA encoding novel
c	32	75	1.6	1551	23	AAS86140	DNA encoding novel
	33	75	1.6	1563	23	AAS84054	DNA encoding novel
34	75	1.6	1584	23	AAS68494	DNA encoding novel	
35	75	1.6	1629	23	AAS85800	DNA encoding novel	
c	36	75	1.6	1629	23	AAS86022	DNA encoding novel
	37	75	1.6	1629	23	AAS87359	DNA encoding novel
c	38	75	1.6	1674	23	AAS86041	DNA encoding novel
	39	75	1.6	1690	23	AAS85792	DNA encoding novel
40	75	1.6	1690	23	AAS86019	DNA encoding novel	
41	75	1.6	1690	23	AAS87350	DNA encoding novel	
42	75	1.6	1810	23	AAS85801	DNA encoding novel	
43	75	1.6	1810	23	AAS86023	DNA encoding novel	
44	75	1.6	1827	23	AAS84506	DNA encoding novel	
45	75	1.6	1827	23	AAS85069	DNA encoding novel	

ALIGNMENTS

RESULT 1
NAAS85786
ID NAAS85786 standard; cDNA; 2079 BP.
AC NAAS85786;
XX
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #21590.
XX
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
XX
PN WO200175067-A2.
XX
XX
PD 11-OCT-2001.
XX
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG21599.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

Query Match 1.9%; Score 88.6; DB 23; Length 2079;
 Best Local Similarity 48.9%; Pred. No. 5.9e-09;
 Matches 331; Conservative 0; Mismatches 334; Indels 12; Gaps 3;

QY 1976 gtatctaattacatgggttatcaactagttattgctatgatgggtattccaaattcgcca 2035
 Db 1827 GTAGCTGAATACACGGCTATCATCGTTGTTGCTCTGCGGAAACCCGAAAGACGCTG 1768
 QY 2036 actggagataatcatatctctttcacacaagtatggtgataagacttttctcat 2095
 Db 1767 ATGACATCAATCTACATGTTTATCAAAAGCTGGAAGACGGCGGCTCTTTAAA 1708
 QY 2096 tggcgcaatgcaggttcaatttttggaaactaaagaacaaatgtgttccagaagtgttca 2155
 Db 1707 GACAGCGATAGTTCGACGCCAACGATCCGATCCTGAAAGATCAGACGCAAGAATGGTCC 1648
 QY 2156 ggttcagctattgataatggtgataaattcaactatttttcaactcaaatgatgacg 2215
 Db 1647 GGTTCGCAACTTTACATCTGACGGAATAATCCGTTTATTCTACAC-----TGACTAT 1594
 QY 2216 tctgattacaagtgtgaatgacacgccttgcacacgaacattaaaccttaattgtgat 2275
 Db 1593 TCCGGTAAACATTACGGCAACAAAGCTGACACAGCGCAGGTAATGTGTCAAATCT 1534
 QY 2276 gataacggtgttcaatcaagaagtgttgataaattatcaagttttgttgaagggtgagga 2335
 Db 1533 GATGACACACTCAAAATCAACGGAGTGAAGATCACAAAACGATTTTTCGCGGAGACGGA 1474
 QY 2336 tttcactaccaacttatgaacattgcaaacgcaacgcaagatcgtgaaaatgatgattac 2395
 Db 1473 AAACATATACAGAGTTACAGCTTTATCGATGAAGCAATTATACATCCGGCGACAC 1414
 QY 2396 tgcttacgtgaccacacgtgtttcaattagaataatggtgatcgcttctgtatttcgaa 2455
 Db 1413 CATACGCTGACAGACCTCTCACTACGTTGAAGACAAAGGCCATAATACCTTGTATTGAA 1354
 QY 2456 gctaatactgggacagaa---gattacaaagtgcagacacaaatataattggtggctaac 2512
 Db 1353 GCCAACACGGGAACAGAAACGGATACCAAGCGCGAAGAAATCTTTTATTAAACAAAGCGGTAC 1294
 QY 2513 tatggtggcgtatgctcttcaatattaaagagttcccttcaagcttttgaataataagaag 2572
 Db 1293 TACGGCGCGCGACGAACCTTCTCCGTAAGAAAGCCAGAGACTTCAGCAGCGCTAAA 1234
 QY 2573 gatcgtgaattggtggttagtctaattggtcacttggttatctttaaagctcactaacaat 2632
 Db 1233 AAACGCGA---TGCTGAGTTAGCGAAGCGCGCCCTCGGTATCATAGATTAAATAATGAT 1177
 QY 2633 caaagtaagcacaaggt 2649
 Db 1176 TACACATTGAAAAAGT 1160

RESULT 3
 ID AAS87342
 XX AAS87342 standard; cDNA; 2079 BP.
 AC AAS87342;
 XX AAS87342;
 DT 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #23146.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX
 XX W0200175067-A2.
 PN
 XX
 PD 11-OCT-2001.
 XX

30-MAR-2001; 200IWO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 DR
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG23155.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1: SEQ ID No 23146; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2079 BP; 710 A; 433 C; 475 G; 461 T; 0 other;

Query Match 1.9%; Score 88.6; DB 23; Length 2079;
 Best Local Similarity 48.9%; Pred. No. 5.9e-09;
 Matches 331; Conservative 0; Mismatches 334; Indels 12; Gaps 3;

QY 1976 gtatctaattacatgggttatcaactagttattgctatgatgggtattccaaattcgcca 2035
 Db 253 gtatgtaatacaacggctacgtgtgtgttcttgcgggaacccgcaagacgctg 312
 QY 2036 actggagataatcatatctctttcacacaagtatggtgataagactttctcat 2095
 Db 313 atgacacataatctacatgtttttacaaaagctggaaaacgcggcgctgtctttaa 372
 QY 2096 tggcgcaatgcaggttcaatttttggaaactaaagaacaaatgtgttccagaagtgttca 2155
 Db 373 gacagcgataagttcgcagcgaacgatccgatcctgaaagatcacgacgaagaatggtcc 432
 QY 2156 ggttcagctattgataatgatggtgataaattcaactatttttcaactcaaatgatgacg 2215
 Db 433 ggttctgaacctttacatctgacgggaaaaatccgtttattctcac-----tgactat 486
 QY 2216 tctgattacaagtgtgaatgatcaacgccttgcacacgaacattaaaccttaattgtgat 2275
 Db 487 tccggttaaacattacgcaaaaacagcctgacacacgcgcaggttaaatgtgtcaaaatct 546
 QY 2276 gataacggtgttcaatcaagaagtgttgataaattatcaagttttgttgaagggtgagga 2335
 Db 547 gatgacacactcaaaatcaacggagtggaagatcacaaaacgatttttgcgagacgga 606
 QY 2336 tttcactaccaacttatgaacaatttcacacgcgcaagatcgtgaaaaatgatgattac 2395

Db 607 aaacatcatcagacgttccagcagtttatcatgatgaaggaatatacatcatcgcgcgaac 666
 Qy 2396 tgcctacgtgaccacacagctgttcaattagaataatggtgatcttcttattgtattcgaa 2455
 Db 667 catcagctgagagaccctcactacgttgtagacacaaagccataatcacctgtattcgaa 726
 Qy 2456 gctaaactactgggacagaa---gattacaaagtgcagacacaaatttataatgggctaac 2512
 Db 727 gcaacacaggggaacagaaacaggtaccacgaagcgagaaatcttattacaaagcgtac 786
 Qy 2513 tatggtggcgatgatgccttcaataattaagagttccttcaagcttttgaataataagaag 2572
 Db 787 tacggcgcgccagcaactcttccgtataaagaacgaagcttcagcagcgctaaa 846
 Qy 2573 gatcgtgaattggctgtttagctaatgttgcactggttgatcttcaagctcactaacaat 2632
 Db 847 aaacgcga---tgctgagttagcgaacgcgcctcggtatcatagagttataataatgat 903
 Qy 2633 caaagtaagccaaaggt 2649
 Db 904 tacacattgaaaaaagt 920

RESULT 4

AAS88703/C

ID AAS88703 standard; CDNA; 2079 BP.

AC AAS88703;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #24507.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.

WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PSDB; ABG24516.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID No 24507; 103pp; English.

The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 2079 BP; 461 A; 475 C; 433 G; 710 T; 0 other;

Query Match 1.9%; Score 88.6; DB 23; Length 2079;

Best Local Similarity 48.9%; Pred. No. 5.9e-09;

Matches 331; Conservative 0; Mismatches 334; Indels 12; Gaps 3;

QY 1976 gtatctaattacatgggttatcaactagttattgctatgatgggttatccaaattcgcga 2035
 Db 1827 GTAGCTGAATACACGGCTATCACGTTGTGTTGCTCTTGGCGGAACCCGAAAGACGCTG 1768

QY 2036 actggagataatcatatctatcttcaacaagtatggtgataatgacttttctcat 2095
 Db 1767 ATGACACATCAATCTACATGTTTATCAAAAGCTGGAACACGGCGCCGTGCTTTAAA 1708

QY 2096 tggcgcaatgcaggttcaatcttggaaactaaagaacaataatgtttccaaagaatgtca 2155
 Db 1707 GACAGCGATAAGTTCGACGCCAACGATCCGATCTGAAAGATCAGACGCAAGATGTGTC 1648

QY 2156 ggttcagctattgaaatgatggtgataattcaactattttcaactcaaatgatac 2215
 Db 1647 GGTTCGCAACCTTTACATCTGACGGAATAATCCGTTTATTCTTACAC-----TGACTAT 1594

QY 2216 tctgattacaagtgaatgatcaacgccttgctaccgcaacattataaacttaattgtat 2275
 Db 1593 TCCGGTAAACATTACGGCAACAAAGCCTGCACACACGCGCAGGTAAATGTGTCAAAATCT 1534

QY 2276 gatacgggtgttcaatcaagagtggtgataattatcaagttttgttgaaggtgagga 2335
 Db 1533 GATGACACACTCAAAATCAACGGAGTGGAAAGATCACAAAAACGATTTTTGACGGAGACGA 1474

QY 2336 ttccactaccaactatgaacattcgcaacggcgaagatcgtgaaatgatgattac 2395
 Db 1473 AAACATATCAGAACGTTTCAGCAGTTTATCGATGAAGGCAATTATACATCCGCGACAAC 1414

QY 2396 tgcttacgtgacccacacgcttgctcaattagaaaatggtgattcttctgtattcgaa 2455
 Db 1413 CATACGCTGAGAGACCCCTCACTACGTTGAAGACAAAGGCCATAATACCTTGTTATTCGAA 1354

QY 2456 gctaatactgggacagaa---gattaccaaaagtgaagaccaaatttataattgggctaac 2512
 Db 1353 GCCAACACGGGAACAGAAAAACGGATACCAAGGGGGAAGAAATCTTTATTAAACAAAGCGTAC 1294

QY 2513 tatggtggcgatgatgccttcaatattagaagttcctcaagcttttgaataataagaag 2572
 Db 1293 TACGGCGCGGCGACGAACTTTCTCCGTTAAAGAAAGCGCAGAAAGCTTCAGCAGCGCTAAA 1234

QY 2573 gatcgtgaattggctgtttagcttaattggtcacttggttcttcttaagctcactaacaat 2632
 Db 1233 AAACGGCA---TGCTGAGTTAGGCAACGGCGCCCTCGGTATCATAGATTAAATATGAT 1177

QY 2633 caaagttaagccaaaggt 2649
 Db 1176 TACACATTGAAAAAAGT 1160

RESULT 5

AAA70105

ID AAA70105 standard; DNA; 5940 BP.

XX

AC AAA70105;

XX

DT 07-NOV-2000 (first entry)
 DE Plasmidium falciparum chromosome 2 related DNA sequence SEQ ID NO:238.
 XX
 KW Plasmidium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoa; infection; insecticide; ds.
 XX
 OS Plasmidium falciparum.
 XX
 PN WO200025728-A2.
 PD
 XX 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 PA (HOFFMAN S. .
 PA (CARUCCI D.
 PA (GARDNER M.
 PA (VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX
 DR WPI; 2000-365347/31.
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmidium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection .
 XX
 PS Disclosure; Page 460-462; 577pp; English.
 XX
 CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmidium falciparum.
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmidium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AA700078 to AA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX
 SQ Sequence 5940 BP; 3106 A; 343 C; 879 G; 1612 T; 0 other;

Query Match
 Best Local Similarity 1.7%; Score 80.8; DB 21; Length 5940;
 Matches 221; Conservative 0; Mismatches 212; Indels 3; Gaps 1;

Qy 1383 atggaataattattagtcgaactaagtaatacaacacagtcgctgcaatacaagaatacat 1442
 Db 125 ataaataataatagaactaaggaagaaagtaaaactaaataataataataataattct 184
 Qy 1443 ctgctactaacaactgaaatgctgtgttaaggaagcaaaaataactaacaatactg 1502
 Db 185 tagatgataataatgaagagacataataataataataataataataataataata 244
 Qy 1503 aaaatgctgtgttaataagaacaaaataactaacaataactgaaatgctgtgttaag 1562
 Db 245 ataatgataacaataatgataataataataataataataataataataataataata 304

Qy 1563 aaacacaaataactaacacacagaaaaacgataatagtcac---attaaagttaactaata 1619
 Db 305 ataatgagaataataatgataataataatttaataataatcagcgatgaaatacaaaa 364
 Qy 1620 atgaacaaccatcagcgcctactcaagcaacttgaagaagctaaatcctcaagctgcta 1679
 Db 365 atattatacataaagacaatagcctagagaacacagcttaagatgatacattaaagtcacatta 424
 Qy 1680 aggcgtgtccaaatgccaaagattgacgcgttagtttaacacagatgacacaaataatgaat 1739
 Db 425 gttcgtgtcgaaataaaatttgaaattacgaagaagtaaaattgaagaattagaaaaagaat 484
 Qy 1740 taaataagattaaacttctctaagtcgtcgtgaaggggtgcaaaattgaacctttaaggact 1799
 Db 485 taaaagaagtaaggaataagataattgataataataatgattgataataataataataaagaaa 544
 Qy 1800 tagaggggattggttaa 1815
 Db 545 aagaagattttgttaa 560

RESULT 6
 AAS86050
 ID AAS86050 standard; cDNA; 3947 BP.
 XX
 AC AAS86050;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #21854.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG21863.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity .
 XX
 PS Claim 1; SEQ ID No 21854; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (II) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in

Qy 2549 tccaagcttttgaataataagaagaagatcgtgaattggtgctgttttagctaatagggtgcact 2608
 |||||
 Db 896 CAGAAGCTTCAGCAGAGCGCTAAAAAAGCGCA ---TGTGATTAGCGAAGCGGCGCTC 840
 |||||
 Qy 2609 ggtatttaagctcactaacatcaaatgaagtgaagccaaagg 2648
 |||||
 Db 839 GGATCATAGAGTTAAATTAATGATTACACATTTGACAAAG 800
 |||||

RESULT 8
AAS86024/c
ID AAS86024 standard; cDNA; 979 BP.

New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity
 Claim 1: SEQ ID No 21828; 103pp; English.
 The invention relates to isolated polynucleotide (I) and
 polypeptide (II) sequences. (I) is useful as hybridisation probes,
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 and gene mapping, and in recombinant production of (II). The
 polynucleotides are also used in diagnostics as expressed sequence tags
 for identifying expressed genes. (I) is useful in gene therapy techniques
 to restore normal activity of (II) or to treat disease states involving
 (II). (II) is useful for generating antibodies against it, detecting or
 quantitating a polypeptide in tissue, as molecular weight markers and as
 a food supplement. (II) and its binding partners are useful in medical
 imaging of sites expressing (II). (I) and (II) are useful for treating
 disorders involving aberrant protein expression or biological activity.
 The polypeptide and polynucleotide sequences have applications in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits to assess biodiversity
 and to produce other types of data and products dependent on DNA and
 amino acid sequences. AAS64197-AAS94564 represent novel human
 diagnostic coding sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

	Query Match	1.7%	Score 76.6;	DB 23;	Length 979;
	Best Local Similarity 50.9%;	Pred. No. 1.6e-06;			
	Matches 265; Conservative 0;	Mismatches 244;	Indels 12;	Gaps	
Qy	2132	acaaatggttccaagaatggtcaggttcagctattgttaaatgtagtggtacaattcaa	2191		
Db	591	ANAGTCAGACGCAAGAATGGTCGGGTTCTGCAACTTTTACATCTGACGGGAAAAATCCGT	532		
Qy	2192	ctattttccacctcaaatgatacgtctgattacaagttgaatgatacaacgctgtctacc	2251		
Db	531	TTATTCTACAC-----TGACTATTCCGGTAAACATTACGGCGCAACAAGGCTGACAACA	478		
Qy	2252	gcaacattaacacttaattgttgatgataacggtgtttccaatcaagagtggttgataattat	2311		
Db	477	GCGCAGGTAAATGTGCAAAATCTGATGACACACTCAAAATCAACGGAGTGGAGATCAC	418		
Qy	2312	caagtttgttgaagtgatgattcactaccacaaacttatgaacaattcgcacaaagcgc	2371		
Db	417	AAACAGATTTTTCACGGGACGGGAAAAACATATCAGAACGTTTCACACAGTTTATCGATGAA	358		
Qy	2372	aagatcgtgaaatgatgattactgtctcacgtgaccacacgctgtttcaattagaaat	2431		
Db	357	GGCAATTATACATCCGGCGNACACCATACGCTGAGAGACCCCTCACTAGTTGAGACAAA	298		
Qy	2432	ggtgatcgttatcttgttattcgaagctaatactgggacagaa---gattaccaaaagtgc	2488		
Db	297	GGCCATAAATACCTTGTATTGGAAGCCAACACGGGAACAGAAACGGATACCAAGGGGAA	238		
Qy	2489	gaccaaatttaattgggctaactatggtgcgatgatgccttcaattataaagagttcc	2548		
Db	237	GAATCTTTATTTAACAAGCGTACTACGGCGGCGCGACGAACCTTCTCCGTAAAGAAAGC	178		
Qy	2549	tccaagcttttgaataataagaaggatcgtgaattggcgtggtttagctaatggtgcactt	2608		
Db	177	CAGAAGCTTCAGCAGACGCGCTAAAAAACCGCA---TGCTGAGTTACGACAGCGGCCCTC	121		
Qy	2609	ggtatcttaagctcactacaatcaaatgaagccaaaggt	2649		
Db	120	GGTATCATAGAGTTAAATATGATTACATTTGAAAAAAGT	80		

RESULT	9
AAF26289/c	
ID	AAF26289 standard; DNA; 5340 BP.
XX	
XX	AAF26289;
XX	
XX	
DT	02-MAY-2001 (first entry)
XX	
DE	P. falciparum eg5 DNA SEQ ID 21.
XX	
KW	Human; eg5; antisense oligonucleotide; egg expression; inhibitor; ds.
XX	
OS	Plasmodium falciparum.
XX	
PN	WO200107602-A2.
XX	
PD	01-FEB-2001.
XX	
PF	21-JUL-2000; 2000WO-EP07345.
XX	
PR	28-JUL-1999; 99DE-1035303.
XX	
PA	(AVET) AVENTIS PHARMA DEUT GMBH.
XX	
PI	Uhlmann E, Greiner B, Unger E, Gothe G, Schwerdel M;
XX	
DR	WPI; 2001-159720/16.
XX	
PT	Antisense oligonucleotide for inhibiting human egg expression,
PT	comprises a sequence corresponding to a nucleotide sequence encoding a
PT	defined part of a human eg5 or its mutated form -
XX	

/note="unnamed protein product"

/codon_start=1

/transl_table=11

/protein_id="CAD19335.1"

/db_xref="GI:17899732"

/translation="MYKVGKNVAATLVLSASILMGVVVTAHADQVESNNYNGVAEYNT
ERQANGIGVDGKIIISANSNTTSGSTNOESSATNTNNTAVVNESKNTNNTNENAVVNE
KNTNTNENAVVNEKNTNNTNENDNSOLKLTNNQPSAQAQKLNKLNQPAKAVQNAK
IDAGSLTDQIINELKINFSKSAEKALFFKOLEGIGNAIVKODPOYAIPIYNAKEI
KNPATYTVDAQTKNAHLVDWSPVQDPTGVYNYMGYQLVIAMGIPNPSPTGDN
HYLLNKYKDNDFSHRNAGSIFGKTENVFQWEGSHYIYQYEFANGKDRENDDY
YLKNDRLATLNLNVDNDGSIKSVNYQVLFEGDGHYIYQYEFANGKDRENDDY
CLRDPHVQLENGDLYVEANTGDEYQSDDOIYNWANYGGDDANIKSSFKLLNNK
KRELAGLAGLILKLTNNOSKPKVEEYVPLVTLMACDEVXXKLGDKYILFSVT
RVSRGSDRELAKDNTIVGNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRAT
YSYAVPVAGHPQVOLLITSYMSKNKDFASGEGNTATWAPSLVQINPDPTTVILARATN
QGWVWDDSRNDMLVLKLEGAANSALPGEWGPVWSLNRSPGLGKPKHPQVQP
KIDPDPQSGQNTKNTVPGNDKPKAGKATPDNTNDPDAQPSGQNTNDPDSAOXSGQ
NKNVTPGNEKQKNTDAKQLPQTKNGKSLAGLYAGSLALFLGLAALIEKRHA"

BASE COUNT 1530 a 837 c 897 g 1345 t 25 others

Query Match 99.5%; Score 4612.6; DB 6; Length 4634;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	gttaacaaagacaaaatttatataattcttcaaatataatttcccaactgaagaacata	60
DB	1	gttaacaaagacaaaatttatataattcttcaaatataatttcccaactgaagaacata	60
QY	61	aatgggtacctgtttgatgggaataataattgtgaactaacggcgccacactctttct	120
DB	61	aatgggtacctgtttgatgggaataataattgtgaactaacggcgccacactctttct	120
QY	121	aatgtgctaggatgcataatgatgataattactagatggcggtttttatacataaac	180
DB	121	aatgtgctaggatgcataatgatgataattactagatggcggtttttatacataaac	180
QY	181	tcgcagagagaagaagaacaaatagtaggaactgcgtactgtatttgggttaaaaggac	240
DB	181	tcgcagagagaagaagaacaaatagtaggaactgcgtactgtatttgggttaaaaggac	240
QY	241	tatttaggtgtaaaaatttcattcgaactgctaggctactgtatttgggttaaaaggac	300
DB	241	tatttaggtgtaaaaatttcattcgaactgctaggctactgtatttgggttaaaaggac	300
QY	301	gcaaacccggaagatccatcaatcccatatttatactggatagaatacaataaagtacg	360
DB	301	gcaaacccggaagatccatcaatcccatatttatactggatagaatacaataaagtacg	360
QY	361	gaattcagatcgagcaaaaaagtaagaatatacaaaaaatatacaaaacggcccatgcta	420
DB	361	gaattcagatcgagcaaaaaagtaagaatatacaaaaaatatacaaaacggcccatgcta	420
QY	421	acaacagacttgcctattatagccacgagtacatacagcgatgctgatagtcacgtgca	480
DB	421	acaacagacttgcctattatagccacgagtacatacagcgatgctgatagtcacgtgca	480
QY	481	acatcatcgcaacgtagagaatatctccaattggatattttcagaagaagaatatgttg	540
DB	481	acatcatcgcaacgtagagaatatctccaattggatattttcagaagaagaatatgttg	540
QY	541	tcttattcactaatgtcattcccccaatgctagtaatttttaatttcgggtattctagaatct	600
DB	541	tcttattcactaatgtcattcccccaatgctagtaatttttaatttcgggtattctagaatct	600
QY	601	aaagtccaactatttgggttaaaaaacttttcggtcggttggaagtcgcattatcggtat	660
DB	601	aaagtccaactatttgggttaaaaaacttttcggtcggttggaagtcgcattatcggtat	660
QY	661	tcaaacactattatttataatttcccttggcgactgttggtagaagccaggamca	720
DB	661	tcaaacactattatttataatttcccttggcgactgttggtagaagccaggamca	720


```

Db 3961 GSCCATGGTATTTCATGGTCTCTTCCTCTGTTGAATAGGATCTGGCCAAATAGA 4020
QY 4021 tagtactcttaacgctoctoctgaataatcattcaagccaaagaaattcacgcccattgatctg 4080
Db 4021 TAGTACTCTCTAAACGCTCTCGAATATCATTCGAAGCCAAAGAAATTCACGCCCATGATCTG 4080
QY 4081 gagtcaatgaatgacaaaattctttagaataagacacctaaagagatacaattaagccctgat 4140
Db 4081 GAGTCATGATGACCAAAATCTTTAGGAATAGACCTTAAGAGATCAATTAAGCCCTGAT 4140
QY 4141 attgaattcggagaaggagttgtccacaattgcccgttataataaccagggttaataac 4200
Db 4141 ATTTGAATTCGGAGAAGGGAGTTGTCCAACAATTCGCGTTATTAATACAGGGTTAATAC 4200
QY 4201 ggccttgggctctacaggttaattgtatttggctcagatcagtgatagaaccacaca 4260
Db 4201 GGCCCTGGGCTCTACGGTAATATTGTATTTTGGCTCAGATCAGTGATAGAAACCCACA 4260
QY 4261 gatttagcttggcgggtgagtgctgagtcgaagtcctcaattacttctgtaaccattgtga 4320
Db 4261 GATTAGCTTGGCGGTGGAGTGTCTTGGAAGCTTCAATTAATCTCGTTACCATGTTGA 4320
QY 4321 ttgctaactgtatgtctgttgggtgtagtaggcataccacacctccataaaa 4380
Db 4321 TTGCTAATCTGATGTGCTGTGTTGGTGTAGTAGCATCATACCACCTCTCATAAA 4380
QY 4381 taaggataacaggaattcttctactatataatgatcattcccaataataatattagccg 4440
Db 4381 TAAGGTATAACAGGAATTTCTTGATATATGATCTTCCAATATATATATTAGGCCG 4440
QY 4441 ataagaaatgaccagctaccattcttctgtagttagtaataataatcagatgatactca 4500
Db 4441 ATAAGAAATGACCAGTACCATTCTTGATGCTTAGTGAATATATCGGATGATAGTCA 4500
QY 4501 cccctcaacaatccatttcacggaggtgagtaatacatcggagagctagggaattgg 4560
Db 4501 CCCCTCAACAATCCAAATTTACGGAGGTGAGTAATCATCGGAGAGCTAGGAATGATGG 4560
QY 4561 aggaacgaacaggtccatcgccagctgctatttggatttagccaaagcagcgttact 4620
Db 4561 AGGAACGAACAGGTCCATCGGCAGTGGCTATTGGATTGATTTAGCCAAAGACGCTTACT 4620
QY 4621 gcttgcaaaagctt 4634
Db 4621 GCTTGCAAAAGCTT 4634

```

```

RESULT 2
LOCUS AX316614
DEFINITION Sequence 2 from Patent WO0190319.
ACCESSION AX316614
VERSION AX316614.1 GI:17899728
KEYWORDS
SOURCE
ORGANISM Lactobacillus reuteri.
REFERENCE Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
AUTHORS van Geel-Schutten,G.H., Rahaoui,H., Dijkhuizen,L. and van
Hijum,S.A.
TITLE Novel fructosyltransferases
JOURNAL Patent: WO 0190319-A 2 29-NOV-2001;
FEATURES
source Location/Qualifiers
1. 2367
/organism="Lactobacillus reuteri"
/db_xref="taxon:1598"
BASE COUNT 814 a 433 c 459 g 661 t
ORIGIN

```

```

Query Match 13.7%; Score 637; DB 6; Length 2367;
Best Local Similarity 59.0%; Pred. No. 3.5e-99;
Matches 1266; Conservative 0; Mismatches 821; Indels 60; Gaps 8;

QY 1273 tttaatggaggggttgtaacgcgtcatgctgaatagaagaatacaaatatacaacgg 1332
Db 114 TTCTACTGTACAAGTTTACAACAGAGGTGATAATGATTCTGTTAAAGTGTGACACTTGG 173
QY 1333 tgttgctgaagttaataactgaacgtcaagctcaatggctcaaatggcgtagatgaaaaaat 1392
Db 174 TAGTGGTCAAGTTAGTGCAGCTAGTGATACGACTATTAGAACTTCTGCTAATGCAAAATAG 233
QY 1393 tattagtgtaacagtaataacaacccagtggtcgacaaatcaagaatcatctgctactaa 1452
Db 234 TGCTTCTTCTGCCGCTAATACACAAAATTCCTAACAGTCAAGTAGCAAGTTCTGCTGCAAT 293
QY 1453 ca----atactgaaaaatgctgtgtgttaa tgaagcaaaaaataactaacaatactgaaaaatgc 1509
Db 294 AACATCATCTACAAGTTCCGCGAGCTTCATTAAATAACACAGATAGTAAGCGGCTCAAGA 353
QY 1510 tgttgtaataagaacaaataactaacaatactgaaaaatgctgttgaatgaaaaaa 1569
Db 354 AAATACTAATACACCCAAAATGATGACGCCAAAAGCTGCACCAGCTAAGCAATCTTC 413
QY 1570 aataactaacaacacagaaacagataa---tagtcaattaaagttaactaataataagaac 1626
Db 414 TGAAGCTAAAATAATGAACACAGCTGTAAACGGTTAATGATTCTTCAGCTGCAAAATAATGATGA 473
QY 1627 accatcagcgcctactcaagcaaaactgaagaagctaaatcctcaagctgtaaggctgt 1686
Db 474 TCAACAATCCAGTAAAAGAAATATACCGGTAAAGTTAAACAAGGATGCTGAAAACGTTGT 533
QY 1687 tcaaaatgcgaagttagtgccggtagtttaacagatgatacaaatgaatgaataataa 1746
Db 534 AAAAAAGCGGGAATTCATCTTAACAGTTTAACGTGATGACCAGATTAAAGCATTTAATAAA 593
QY 1747 gattaactcttaagctctgctgaaaaaggttgcaaaattgaccttaaggacttagaggg 1806
Db 594 GATGAACCTTCGAAAAGCTGCAAGTCTGTGTACACAATGACTTATATGATTATGTTCCAAA 653
QY 1807 gattggttaatgctattgttaagcaagatcccaaatatgctattccttactcctaagctaa 1866
Db 654 GATTGCTGATACGTTAATCAACAAGATGTCGTCGACACAGTTCCATCTTTAANGCAAG 713
QY 1867 ggaatacaagaatatgcctgcaacatacacgttagatgccccaaacaggtgaagtggtca 1926
Db 714 TGAATCAAAAATATGCTGCCGCTACAACTAAAGATGCACAAACTAATACTATTGAACC 773
QY 1927 tcttgatgctgggactcttgccagtcacaagatccctgacaggttatgtatataatta 1986
Db 774 TTTAGATGATGGGATTCATGGCCAGTTCAAGATGTTTCGGACAGGACAAAGTGTCTAATTG 833
QY 1987 catgggttatcaactagttattgctatgaggtggttattcccaaatcgcaactggagataa 2046
Db 834 GAATGCTATCAACTTGTCTATCGCAATGATGGGAATTTCCAA-----CCAAAATGATAA 887
QY 2047 tcaatactcttcttatacaacaagtatgggtgataatgactttcttoatttgcgcgaatgc 2106
Db 888 TCATATCTATCTCTTATATAAATGAATGCTGATGATGAATTAAGTCATTTGGAAGAATGT 947
QY 2107 aggttcaactcttggaaactaaagaaacaaatggttccagaatgctcaggttcaagctat 2166
Db 948 AGGTCCAATTTTGGCTATAATTTCTACCGCGGTTTCACAGAAGATGGTCAGGATCAGCTGT 1007
QY 2167 tgtaaatgatgagtggtacaaattcaactattttcaacctcaaatgatacgtctgattacaa 2226
Db 1008 TTTGAACAGTGTGATCACTCTATCCAAATTTATTAACAAGGGTAGACACGCTCTGATAACAA 1067
QY 2227 gttgaatgatcaacgcttgcctacccaacattaaaccttaattgttgatgataacgggtgt 2286
Db 1068 TACCAATCATCAAAAATTTGCTAGCGCTACTCTTTTATTTAACTGATAATAATGAAATGT 1127
QY 2287 ttcaatcaagagtggttgataattatcaagtttggttgaaggtgagtgatttcaactacca 2346

```


similar to Chlamydia trachomatis Hom1: SwissProt Accession
Number P55137"
/citation=[2]
/codon_start=1
/transl_table=11
/product="YefA"
/protein_id="AAB66480.1"
/db_xref="GI:2318067"
/translation="AKRLRRSPADNCGDEAAGVCVL"
complement(227..1140)
/gene="yefB"
complement(227..1129)
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="YefB"
/protein_id="AAB66472.1"
/db_xref="GI:2318059"

gene

CDS

/translation="MYNEGVTSQARKKATDKWHGSSIRKILLENPHYGTLOOQR
EYRSVTSKRRSVNLNQLIIIEKSHPEIIPLEDPLIVQELITRKRPPQAEVHLFT
NTAVKDCRSRHHKRRKGYCSYKNGHGLKACSDHVTSEIDLTDRVITNLNKLK
FSKNYKELSEKALKYKEKTESKIIIEINELNDKDKSNLIVISLANGVSKEDYQL
AINITNEDISNFTLHLSKOLEFKIEKIIDFKSLDKFMKEGTLTPMLHLVD
EIEYHANGTIEINRYFREPTVPSA"
1024..1445
/gene="yefC"
1024..1030
/gene="yefC"
1068..1445
/gene="yefC"
/note="similar to Enterococcus faecalis plasmid PY117 orf4
encoded by GenBank Accession Number D78257"
/citation=[2]
/codon_start=1
/transl_table=11
/product="YefC"
/protein_id="AAB66473.1"
/db_xref="GI:2318060"

gene

RBS

CDS

/translation="MSVAFETLRASCEGLVTPSLNPLAIASKPFPAYISEKILFTIF
SAPAEFRDMIVERTQSGMLAKONPDFRGRPKFTKQINHALTLLENHSYQVED
MTGISVTLVRKKAKAAEAING"
complement(1133..1140)
/gene="yefB"
1505..4159
/gene="yefB"
1505..1512
/gene="yefB"
1520..4159
/gene="yefB"
/note="similar to Methanococcus janaschii plasmid PURB801
hypothetical protein MJBCS02 encoded by GenBank Accession
Number L77119"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefA"
/protein_id="AAB66474.1"
/db_xref="GI:2318061"

RBS

gene

RBS

CDS

/translation="MALIDEDKIAEIVNRHSDFLYELLYGVVDVPRATITRUKGN
QNTKRVGHLKRWKFKKGLDFDALIDEEQVEYLSAKPRYLIVTDYGLAK
DTITLALDYFELPYDFDFLAWGIERFEKENPADIKAAERFARIYDVLRKEN
NIETNRGLDFLIRLFCFAEDTDFIKRNSFTNLITLTDEGSLNKLFAFLIV
LDKNERDVPYLVNGQLFTPEPTLEFSAKSRKLIIECGELLNWKINPDLF
GSMIOAVPESRSYGLMHVTSVPMKVIKPLFLDKLNSFLDAYDDYTKLENLLTR
IGKIKDFPACGSGNELIYIKELRHEINIRKLQELGELIYVPSVTSQFYGIEI
EDAHVAKULSWIAEHMNEELKRVNHNVRPLHTAGDIRCANARIVETVECP
AQGSEEVYFNGPYLGSQKQKHKSDMLSIFCYKNGKMLDYSIAWFYFGAKYAST
TNKAFVTSNVSQGVSLWNLKFKGICQINFAKSKWANNKNAAVIVIVIG
FGPLDKVNYLVDETKLVSNISPYLTDGENTLVSSRTKPISDLPLHGNPNDOG
GGLFTITEYTDANKYPELVYKPKFVIGSVFINGLRLCLNLAKYKIKPNPLI
QERISIKNRHRSKDGTKNLALTTPWKFRDTHETTNYSIVVPSVSSSENRIYIPMGLA
GADTILSLNLIYIDAEIYLLGLMSRHMHTWYKAVAGRLKTDYRSAGLCYNTFPPI

RBS

gene

CDS

EISTRKRKTEIEAILEILDLREQGGTAEALYNPSTWPIELKVAHEKLDGIVERAVRQ
KOFESDEERLEVLKLYQEWTER"
4152..4161
/gene="yefB"
4152..5967
/gene="yefB"
4171..5967
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66475.1"
/db_xref="GI:2318062"
/translation="MENIYEINYNOTGSKKKTNEYGMREOMARAFKRNISOYLLVKAP
PASGRSALMFGLDKLIINOGLRKVIVAVPERSIGSEKNTOLKSYGFENKWKYDPRN
NLTVGDSNKSXKVSFVRFMESDDOVLCTHSTLRFAPEKIDDKAFDNCLLAIDEFHVS
ADVNSRLGELLRSIIHNSAHIVAMTGSYFRGDSVFPILLPEDEELFDKVTYSYEQLD
GXEYKGFQIGYHIIYOGQYSAINVEYLDNKKTIIVHIPNVNSESSTKDYDEVKILID
MIGEVEYODDDGTIIYVKRHSQKTLKIALDVDDQVGRNVVAYLRNIEALDLDII
ALGMAKEDPWFCEHTLTGYRGSITFVQIGRTRDSYKNTYAOFTNLAMPDAK
DEVVYTVNTMLKATISALLMEQVLTDPDFKRRRNESEKSTTGELFVKGLKEPSTE
NVKTIENDINDLKAKTCKIHKSKTFSGEVDPKLVNKLVIPIVKVYVNLSEDEIEE
VPOHVLTVMKGAKEVVGSEKFIEMADKFINIEDLDINLIDSNINFPQKAFEVLSKE
LNSPVKLQETIDAKRINFDEDELVFIWPKV"
6064..6069
/gene="yefC"
6064..7228
/gene="yefC"
6080..7228
/gene="yefC"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefC"
/protein_id="AAB66476.1"
/db_xref="GI:2318063"

RBS

gene

CDS

/translation="WASNRHYSINEIMSOLFQYITTPPKTKQAOYDPEVEKFLI
EFVKNRGPQKVPDTLTERSASRLIGIRKDPDRMEYLKDEIGLLEVRSEKELTIP
KISSDDILKSGLSALLGDNLDNITRSIFDTSSLOKVTTPMEYVAKRKKIKDFGKFE
ELFKKHEITEGKRKILTFKNEQDITQNSFYILKGVLLYVEDVGERKKAKCTNARL
RCIPNGTSEMDLRLSALRYKHGRVYTDNEDTLLDNVREDVDVSTGFYIVLKSISTD
PQISSIKLYKIGFTTGSVENIRNAENQSTLYAPVEIVTIVQFMNNAKSFATAIH
HALENNLDVSLGANGKMLVPKPFVVTLEDLQAVIDEIVMMVHLVD"
complement(7490..8180)
/gene="yefB"
complement(7490..7795)
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66477.1"
/db_xref="GI:2318064"
/translation="MTVRKELIKOINLTIVTKTINOKNPTPMVKNILKRYEAEKEFI
LQSSDGKEEDLSKVRNKLDTLTRYLESANDYMPMLKEMHKTEKLLKEYDETTOF"
complement(8175..8180)
/gene="yefB"
8277..8298
/note="rho independent terminator; deltaG= -10.2kcal/mol"
/citation=[2]
complement(8308..9470)
/gene="yefB"
complement(8308..9417)
/gene="yefB"
/citation=[2]
/codon_start=1
/transl_table=11
/product="yefB"
/protein_id="AAB66478.1"
/db_xref="GI:2318065"

gene

CDS

RBS

terminator

gene

CDS

/translation="MPAGVPYNAVNVGTLKEGLQKFAKVLPKPYGTSSSGRRTPAPH
VPPVTVKYGEHFAWRKKVKLPNLIYKTEGTYTTDNYGRITSVKADQLGCEAKRN

[illegible]

us-09-995-587a-10.rge

[illegible]

Search completed: September 22, 2002, 05:40:23
Job time: 21803 sec

Db 55 -OENTTA-----SSAAENQAKTEYQETPSTNPAATAVENTDOTTKVITDAAVESKASK 108
Qy 114 --NENKNTNTENDNSOLKLTNNEOPSAT--QANLKK-----LNPQAAKAVONAKI 161
Db 109 TKDQATVTKAATPEVQGTQNEKDKAKATEADITTPKNTIDEYGLTEQARKIATEAGI 168
Qy 162 DAGSLTDQINELNKNFNSKAEKAKITFKDLEGIGNAIVKQDPOYAIKPSNAKEIKNM 221
Db 169 NLSSLTQKOEALNKVLTSDAQTHQMTYQEFKIAOTLTAODERYAIPFNKAIKNM 228
Qy 222 PATYTVDAQTKMAHLVDWSPVQDPVTGYSVSNVMGQVLVIAMGIPNSPTGDNHIIYLL 281
Db 229 KAATTRAQTOGIAADLDVWDSWPQDAKTGEVINWNGYQLVVAMGIPN--TNDNHIIYLL 286
Qy 282 YNKYGDNDFSHRNAGSTFGTKETNVFQWGSATVNDGDTIQLFFTSNDTSDYKLNQDR 341
Db 287 YNKYGDNDFHWNKAGSTFGTKETNVFQWGSATVNDGDTIQLFFTSNDTSDYKLNQDR 346
Qy 342 LATATLNVNDNGVSKSVNYQVLF--EG-DGFHYQTYEQFAN---GKRENDDDYCLRD 396
Db 347 LATATVNLGDFDQDVRILSVENDKVLTPGVMAYHYQSYQQWRSTFTGA---DNIAMRD 402
Qy 397 PHVQLENGDRYLVEANTGTEDYQSDDOIYNWANYGGDDAFNFKSSFKLLNNKDRDLA 456
Db 403 PHVIEDENGDRYLVEASTGTENTYQGEDQIYNFTNYGSSAYNKSLFRFLDDQDMYNR 462
Qy 457 GLANGALILKLTNNQSKPKVEEYVSLVSTLMACDEV---XXKLGDKYVLFSTVRSR 512
Db 463 SWANAAGILKLGDKKTPQVQYFTPLLSSTWYSDLELERNVVKLGDKYVLFSTVRSR 522
Qy 513 GSDRELFAKNTIYGVNDVAMTGYVSDSLMGKPKPLNNSGVVLTVASVPANWRTATSYAV 572
Db 523 GSNDAWNAKANEVVDVNVMLGYVSDQLTNGYKPLNNSGVVLTVASVPANWRTATSYAV 582
Qy 573 PVAGHPQVLTYSWNSKDPASGEGNTATWAPSLVQINPDQTTVLARATNQGDWDD 632
Db 583 PVAGSDTLMTATYWRNEVAGKKNSTWAPSLVQINPDQTTVLARATNQGDWDD 642
Qy 633 SSRNDNMLGVLEKGAANSALPGWEGKPKVDMSLNRSFGLKPKHPQVQPKIDPDQDQPS 692
Db 643 PSRTTDTVGTL-----DTAVLPGENDGYIDWNVIG---GYGLKPKHTPGQ-----XQPT 687
Qy 693 GQNT-----KNTVPGNDGKPKAGKATPDNTNTDPSAQPSGQ 727
Db 688 VPSPITHTDDIISVEFGDHLVTKPVKYNDSAGRIDQSRNSGGSLNVAFNVA---GG 744
Qy 728 NNTIDPSAQSGQNTKNTVTRNEKQKNTDAKQLPQTKGKSLAGLAGSLLALF 782
Db 745 NISVKPS-QKSINTKET-----KKAHVSTEEKQKKGNS-----PFAALLALF 787

RESULT 2

SACB_STRSL STANDARD; PRT; 969 AA.
AC Q55242;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
DE transferase) (Sucrose 6-fructosyl transferase).
GN TTF.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxID=1304;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25975;
RC MEDLINE=93322332; PubMed=8331080;
RA Rathsam C., Giffard P.M., Jacques N.A.;
RT "The cell-bound fructosyltransferase of Streptococcus salivarius: the
RT carboxyl terminus specifies attachment in a Streptococcus gordonii
RT model system.";

RL J. Bacteriol. 175:4520-4527(1993).
CC -!- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL) (N) -
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL) (N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -!- SUBCELLULAR LOCATION: SECRETED, CELL-WALL BOUND.
CC -!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L08445; AAA71925.1; -
CC InterPro; IPR003469; Glyco_hydro.68.
CC Pfam; PF02435; Glyco_hydro.68.1
CC Transferase; Glycosyltransferase; Signal; Cell wall.
FT SIGNAL 1 ? POTENTIAL.
FT CHAIN 1 ? 969 LEVANSUCRASE.
SQ SEQUENCE 969 AA; 103983 MW; D389B5B32ACF735A CRC64;

Query Match 38.2%; Score 1586; DB 1; Length 969;
Best Local Similarity 41.6%; Pred. No. 2.8e-77;
Matches 349; Conservative 133; Mismatches 266; Indels 90; Gaps 18;
Qy 2 YKVGKNVATLVLSASILMGVVTVAHADQVBSNNTYNGVAEVNTERQAGQIGVDGKIISA 61
Db 24 YSIRK---VATVGATSVLGLTALFLGATQVRADQVETAPAVATATATPSTASLTAS 80
Qy 62 NSNTTSGSTNQSSTNTNENAV--VNESKNTNTENAVNENKNT-----105
Db 81 ETATSVATSEAVESVAHSEVATKPVETQPSNTTPSYVEEKASSTVTVSSDATTPTSAT 140
Qy 106 -----NNTENAVNENKNTNTNENDNSQLKLTNNEQPSAATQANLKLKPOAAKAV-Q 157
Db 141 VAAVSAPATSEAAVEAPTSTASSEADTHTEVDLKVSENSAANLKLKGRKISYVEE 200
Qy 158 NAKTD-AGSLTDDQINELNKNFNSKAEK--GAKLTFKDLLEGIGNAIVKQDPOYAIYSN 214
Db 201 NMTSDQIYALTEETKALNKYDFSDDAIKGTCTSLTYENLKDIVASFCLKVAPVYFK 260
Qy 215 AKETKNPATYTVDAQTKMAHLVDWDSWPVQDPVTGYSVSNVMGQVLVIAMGIPNSPTG 274
Db 261 ADTINMPAFNTVDAQTKKEIDVWDSWPQDAKSGVYSNWNQYQLVISWAGAPNK--N 318
Qy 275 DNHYLLYKYGDNDFSHRNAGSTFGTKETNVFQWGSATVNDGDTIQLFFTSNDTSD 334
Db 319 SNHYLLYKYGDNDFTHWKNAGPIFGYNALEDQDQWSSSATVNSDGSILQIYTKNDTSG 378
Qy 335 YKLANDORLATATLNVNDNGVSKSVNYQVLFEGDGFHYQTYEQFA-----NGK 385
Db 379 GKLNNQQLASATLNLAVENDEIIVKIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII 438
Qy 386 DRENDYCLDRPHVQLENGDRYLVEANTGTEDYQSDDOIYNWANYGGDDAFNKSSEFK 445
Db 439 PDRDNYCLDRPHIIE--DNGSRYLIFESNTGDNQYQGEKQIYKWSNYGGDDAFNKLSPIN 497
Qy 446 LLNNKKRELKLAGLALGILKLTNNQSKPKVEEYVSLVSTLMACDEV---XXKLGDK 501
Db 498 IVNNKHLYNLASWANGSIGILKLDNNEKNPSVAELYPLVTSHMTDEVERPSVVKMGK 557
Qy 502 YLFSVTRVSRGSDRELTAQNTIVGDNVAMTGYVSDSLMGKPKPLNNSGVVLTVASVPAN 561
Db 558 YLFTSRINKSTDAEGTVAAAREAGDDVVMLEFVSLRGKYPRLNGSGVYLTVASVPAD 617
Qy 562 WRTATSYAVVPVAGHPDQVLTYSWNSKNDKDFASGEGNTATWAPSLVQINPDQTTVLAR 621
Db 618 WRTSTSYAVVPVSGSDTLTLLVTSYMTNRGGIAGAKNKTWAPSLFKMNAADDTTEVLPK 677
Qy 622 ATNOGDVWDDSSRNDNMLGVLEKGAANSALPGWEGKPKVDMSLNRSFGLKPKHPQVQ 681

```

Db 678 MTNCGDWTKSSBS-----LVHVGQNSAKLPNE-DFNVDIYAVS---GYGLKPH--TY 726
Qy 682 PKIQDPQPSQNTKNTPCNG-----DKPAGKATP-----DNTNIDPSAOPSGON 728
Db 727 PTVGSGVSEAHGVLTVVKDKKADKPTVPSPTEGNSVDKTKNPKTSKPADNN 786
Qy 729 TNIDPSAQXSQNTKNTVP-----GNEKGGKNTDAKQLPOTGNKSG 769
Db 787 ---QPSADKEDKPTNPTNPDSPARTPEFYGGHSDNNDNSNDHHVAVPVKPTGDSVG 841

RESULT 3
SACB_BACAM STANDARD; PRT; 472 AA.
ID SACB_BACAM
AC P21130;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transferase) (Sucrose 6-fructosyl transferase).
GN SACB.
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23844;
RX MEDLINE=91092506; PubMed=2265762;
RA Tang L.B., Lenstra R., Borchert T.V., Vasantha N.;
RT "Isolation and characterization of levansucrase-encoding gene from
Bacillus amyloliquefaciens.";
RL Gene 96:89-93(1990).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- INDUCTION: BY SUCROSE.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52988; CAA37179.1; -
DR PIR; JQ0802; JQ0802.
DR PIR; S11739; S11739.
DR InterPro; IPR003469; Glyco_hydro.68.
DR Pfam; PF02435; Glyco_hydro.68; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 472 LEVANSUCRASE.
FT SEQUENCE 472 AA; 52859 MW; F38592D273677E7D CRC64;

Query Match 18.1%; Score 751.5; DB 1; Length 472;
Best Local Similarity 38.0%; Pred. No. 3e-33;
Matches 190; Conservative 88; Mismatches 171; Indels 51; Gaps 19;

Qy 144 NIKKLNPOAKAVQNAKIDAGSLTDQINELNKNFNSKSAEKALTFKDELEGNAIVK 203
Db 2 NIKKIYQATVLTFTTALLAGGATQAFKENQKAY-KETYGVSHITRHDMLQPKQ-Q 58
Qy 204 QDPQAIYPSNAKEIKNMPATYTDQATGKMAHLWDWDSWFPVODPYGVYNTYNGVLVI 263
Db 59 QNEKYQVPQFDQSTIKN-----IESAKG-----LDWDSWPLQN-ADGTVAEYNGYHVF 107
Qy 264 AMGIPNPTGDNHIIYLLNKNYKNDGDFSHRNAGSIFGKTNVF-----QEWS 312

```

```

Db 108 ALAGSPKD-ADDTSTFTYKVGDNDSIDSKNAGRVE--KSDKFDANDPILKQQTQWS 164
Qy 313 GSAIVNDGDTIQLFFTSNDTSYKLANDRLATLNLNDONGVSIKSDVNYQVLFEGDG 372
Db 165 GSATFTSDCKIRLFYT--DYSGKHGKQSLTTAQVNVSKSDDTLKLINGVEDHKTIFDGDG 222
Qy 373 FHYQTYEQFAN-GKRENDYCLRDPHVQLENGDRYLFPEANTGTED-YQSDQIYNWA 430
Db 223 KTYQNVQOQFIDEGNYTSGDNHTLRDPHYVE-DKGHKYLFPEANTGTENGYSQGESLFNKA 281
Qy 431 NYGGDDAFENIKSSFLLNNKKDRELALANGALGILKLTNNQSKPKVEEYVSLVSTLMA 490
Db 282 YGGGTNFFRKESQKLOQSAKKRD-AELANGALGIELNNDYT--LKKYKPLIITSNTV 337
Qy 491 CDEV-----XXKLGDKYLYFSVTRVRSRGSDRELTAKNQNTIVGDNVAMIGYVSDSLMGKYPK 546
Db 338 TDEIERANVFKMGKWLFT--DSRGSKMTI---DGINSNDIYMLGYVSNLSLTGPYKP 390
Qy 547 LNSGVVFLASVPANRTATYIYVAVGHPDOVLITSTWNSKNDPASGEGNTATWAPSF 606
Db 391 LNKTLGLVQLMGLDPNVDVTFYSHFAVQA-KGNVVTISYTNRGEF--EDKKATFGPSE 447
Qy 607 LYQINPDDTTTLARATNOG 626
Db 448 LMNIKGNKTSVVRKNSILEQG 467

RESULT 4
SACB_BACSU STANDARD; PRT; 473 AA.
ID SACB_BACSU
AC P05655; P70984;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transferase) (Sucrose 6-fructosyl transferase).
GN SACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=85295507; PubMed=2993818;
RA Steinmetz M., Le Coq D., Aymerich S., Gonzy-Trebol G., Gay P.;
RT "The DNA sequence of the gene for the secreted Bacillus subtilis
enzyme levansucrase and its genetic control sites.";
RL Mol. Gen. Genet. 200:220-228(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA Denizot F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-62 FROM N.A.
RC STRAIN=168 / P779;
RX MEDLINE=84178454; PubMed=6424671;
RA Fouet A., Arnaud M., Klier A., Rapoport G.;
RT "Characterization of the precursor form of the exocellular
levansucrase from Bacillus subtilis.";
RL Biochem. Biophys. Res. Commun. 119:795-800(1984).
RN [4]
RP SEQUENCE OF 1-68 FROM N.A.
RX MEDLINE=87008406; PubMed=2428811;
RA Shimotsu H., Henner D.J.;
RT "Modulation of Bacillus subtilis levansucrase gene expression by
sucrose and regulation of the steady-state mRNA level by sacD and
sacQ genes.";
RL J. Bacteriol. 168:380-388(1986).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC ALSO ACT AS FRUCTOSYL ACCEPTORS.

```


CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M14202; AAA22725.1; -
 CC EMBL; Z94043; CAB08015.1; -
 CC EMBL; K01987; AAA22724.1; -
 CC EMBL; X02730; CAA26513.1; -
 CC EMBL; Z99121; CAB15450.1; -
 CC PIR; A25040; A25040.
 CC PIR; S07309; S07309.
 CC Subtilisin; BG10388; sacB.
 CC InterPro; IPR003469; Glyco_hydro_68.
 CC Pfam; PF02435; Glyco_hydro_68; 1.
 CC Transferase; Glycosyltransferase; Signal; Complete proteome.
 CC SIGNAL 1 29
 CC CHAIN 30 473 LEVANSUCRASE.
 CC CONFLICT 12 12 V -> I (IN REF. 3).
 CC SEQUENCE 473 AA; 52971 MW; 3FBF2F571B41D5B0 CRC64;

Query Match 18.08; Score 746.5; DB 1; Length 473;
 Best Local Similarity 36.38; Pred. No. 5.5e-32;
 Matches 185; Conservative 92; Mismatches 162; Indels 71; Gaps 18;

Qy 144 NLKLNPOAAKAVQNAKIDAGSLDDQINELNKFSAEKGAKLTKFDLEGIGNAIVK 203
 Db 2 NIKKFAKQATVLTFTTALLAGGATQAFAKETN-----QKPKETYGISH-ITR 48
 Qy 204 QD-----POYAIPIYNAKEIKNNPATYTVDAQTGKMAHLDVWDSWPQDPVTGYV 253
 Db 49 HDMLQIPQOKNEKYQVPEFSSSTIKNISSA-----KGLDVWDSWPLQN-ADGTV 97
 Qy 254 SNYMGYQLVIAAMGIPNSPTGDNHLYLYNKGNDNFDSHWRNAGSIFGTKETNVF----- 308
 Db 98 ANYHGYHIVFALAGDPKN-ADDTSIYMFYQKVGETSIDSWNAGRVF--KDSKDFDANDS 154
 Qy 309 -----QWMSGSAIVNDGTTQLPFTSDTSDYKLNQDLATATLNLVNDNGVSIKSDV 362
 Db 155 ILKQDQTOEWSSGSAFTSDGKIRLFYT--DFSGKHVGKQTLTAAQVNVASDSSLNNGVE 212
 Qy 363 NYQVLFEGDGHYQTYEQFAN-GKDRNDYCLDRPHVVQLENGDRYLVFANTGTED-Y 420
 Db 213 DYKSIFDGDGTYQNVQOFIDEGNYSNGDNHTRDPHYVE-DKGHKYLVFEANTGTEDGY 271
 Qy 421 QSDQIYNWANYGDDAFNIKSSFKLLNNKDKRELALGALGILKLTNNOSKPKVEEV 480
 Db 272 QGEESLFNKAYIGKSTFFRQESOKLLQSDKKR-TAELANGALGMIELNDYTT---LKKV 327
 Qy 481 YSPLVSTLMACDEV-----XXKLGDKYYLFSVTRVSRGSDRELTAKTNTIVGDNVAMIGYV 536
 Db 328 MKPLIASNTVTDETERANVFKNMGKWLFTDSRGSKMTIDGTSND-----IYMLGYV 380
 Qy 537 SDSLMGKYKPLNNSGVVLTASVPANWRTATYVYAVPVAGHPDOVLTSYNSKNDFASGE 596
 Db 381 SNSUTGPKPLNKGKLVKMDLDPNDVFTYSHFVAPQQA-KGNVNVITSYMTNRGFIADK 439
 Qy 597 GNYATWAPSLVQINPDPTTTLARATNOG 626
 Db 440 --QSTFAPSFLNKGKTSVVKDSILEQG 467

RESULT 5
 SACE_BACST
 ID SACE_BACST PRT; 473 AA.
 AC P94468;

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
 DE transferase) (Sucrose 6-fructosyl transferase).
 GN SABC OR SURB.
 OS Bacillus stearothermophilus.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Geobacillus.
 OC NCBI_TaxID=1422;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ATCC 12980;
 RC MEDLINE=98007870; PubMed=9349714;
 RX Li Y., Triccas J.A., Ferenci T.;
 RT "A novel levansucrase-levanase gene cluster in Bacillus
 RT stearothermophilus ATCC12980".
 RL Biochim. Biophys. Acta 1353:203-208 (1997).
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -
 CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
 CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U34874; AAB97111.1; -
 CC InterPro; IPR003469; Glyco_hydro_68.
 CC Pfam; PF02435; Glyco_hydro_68; 1.
 CC Transferase; Glycosyltransferase; Signal.
 CC SIGNAL 1 29 BY SIMILARITY.
 CC CHAIN 30 473 LEVANSUCRASE.
 CC SEQUENCE 473 AA; 52854 MW; 12F7A272EC2B771E CRC64;

Query Match 17.68; Score 730.5; DB 1; Length 473;
 Best Local Similarity 35.98; Pred. No. 3.9e-32;
 Matches 183; Conservative 93; Mismatches 163; Indels 71; Gaps 19;

Qy 144 NLKLNPOAAKAVQNAKIDAGSLDDQINELNKFSAEKGAKLTKFDLEGIGNAIVK 203
 Db 2 NIKKFAKQATVLTFTTALLAGGATQAFAKETN-----QKPKETYGISH-ITR 48
 Qy 204 QD-----POYAIPIYNAKEIKNNPATYTVDAQTGKMAHLDVWDSWPQDPVTGYV 253
 Db 49 HDMLQIPQOKNEKYQVPEFSSSTIKNISSA-----KGLDVWDSWPLQN-ADGTV 97
 Qy 254 SNYMGYQLVIAAMGIPNSPTGDNHLYLYNKGNDNFDSHWRNAGSIFGTKETNVF----- 308
 Db 98 ANYHGYHIVFALAGDPKN-ADDTSIYMFYQKVGETSIDSWKTGPRVF--KDSKDFDANDS 154
 Qy 309 -----QWMSGSAIVNDGTTQLPFTSDTSDYKLNQDLATATLNLVNDNGVSIKSDV 362
 Db 155 ILKQDQTOEWSSGSAFTSDGKIRLFYT--DFSGKHVGKQTLTAAQVNVASDSSLNNGVE 212
 Qy 363 NYQVLFEGDGHYQTYEQFAN-GKDRNDYCLDRPHVVQLENGDRYLVFANTGTED-Y 420
 Db 213 DYKSIFDGDGTYQNVQOFIDEGNYSNGDNHTRDPHYVE-DKGHKYLVFEANTGTEDGY 271
 Qy 421 QSDQIYNWANYGDDAFNIKSSFKLLNNKDKRELALGALGILKLTNNOSKPKVEEV 480
 Db 272 QGEESLFNKAYIGKSTFFRQESOKLLQSDKKR-TAELANGALGMIELNDYTT---LKKV 327
 Qy 481 YSPLVSTLMACDEV-----XXKLGDKYYLFSVTRVSRGSDRELTAKTNTIVGDNVAMIGYV 536
 Db 328 MKPLIASNTVTDETERANVFKNMGKWLFTDSRGSKMTIDGTSND-----DGITSNDIYMLGYV 380

QY 537 SDSLKGYKPLNNGVLTASYPANWRTATSYIYAVPVAGHPDOVLTSYMSKNKDFASGE 596
 DB 381 SNLSGPGYKPLNGLVLMKMDLPNDVYTFYSHFAVPA-Q-TGNVYVITSYMTNRGFIADK 439
 QY 597 GNYATWAPSLVQINPDFTTTVLARATNOG 626
 DB 440 --QSTFAPSLNIOGKTSVVKASILDOG 467
 RESULT 6
 ID INVB_ZYMMO STANDARD; PRT; 413 AA.
 AC 060115; Q60117; Q60125;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-DEC-1998 (Rel. 37, Last sequence update)
 DE Extracellular sucrose (EC 3.2.1.26) (Beta-fructofuranosidase)
 DE (Invertase) (Saccharase) (Protein B46).
 GN SACC OR INVB OR SUCE3.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ATCC 10988 / ZM1;
 RX MEDLINE=94368848; PubMed=8086457;
 RA Song K.B., Lee S.K., Joo H.K., Rhee S.-K.;
 RT "Nucleotide and derived amino acid sequences of an extracellular
 RT sucrose gene (inVB) of Zymomonas mobilis ZM1 (ATCC10988).";
 RL Biochim. Biophys. Acta 1219:163-166(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IFO 13756;
 RX MEDLINE=95218269; PubMed=7766026;
 RA Kyono K., Yanase H., Tonomura K., Kawasaki H., Sakai T.;
 RT "Cloning and characterization of Zymomonas mobilis genes encoding
 RT extracellular levansucrase and invertase.";
 RL Biogsci. Biotechnol. Biochem. 59:289-293(1995).
 RN [3]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-NRRL B806;
 RX MEDLINE=95297907; PubMed=7778976;
 RA Kannan R., Mukundan G., Alt-Abdelkader N., Augier-Magro V.,
 RA Baratt J., Gunasekaran P.;
 RT "Molecular cloning and characterization of the extracellular sucrose
 RT gene (sacc) of Zymomonas mobilis.";
 RL Arch. Microbiol. 163:195-204(1995).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 31821 / ZM4 / CP4;
 RA Ahn J.Y., Kang H.S.;
 RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
 CC fructofuranoside residues in beta-D-fructofuranosides.
 CC -!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AF081588; AAA61488.1; -
 DR EMBL; D17524; BAA04476.1; -
 DR EMBL; L33403; AAC36942.1; -
 DR EMBL; AF313764; AAG29871.1; -
 DR InterPro: IPR003469; Glyco_hydro_68.
 DR Pfam: PF02435; Glyco_hydro_68; 1.
 KW Hydrolase; Glycosidase.
 RN 85
 G -> A (IN REF. 3).
 FT CONFLICT 85

FT CONFLICT 105 105 S -> L (IN REF. 2).
 FT CONFLICT 108 108 T -> S (IN REF. 2).
 FT CONFLICT 404 407 PWP -> LGMA (IN REF. 3).
 SQ SEQUENCE 413 AA; 46100 MW; BACC12D167332B47 CRC64;
 Query Match 7.4%; Score 309; DB 1; Length 413;
 Best Local Similarity 26.5%; Pred. No. 9e-10;
 Matches 105; Conservative 71; Mismatches 150; Indels 70; Gaps 20;
 QY 239 VWDSPVQDPVTGYVSNMGYQVLVIAMGIPNSPTGNHLYLLNKKYGDNDFSHWRNAGS 298
 DB 42 LMDTWPLRD-INGNPVSFKGNVIFSLVADRNPWDRSHARIGYFYSKDGKSWYGGH 100
 QY 299 IFGTKETNFQEWSSAIV--NDGTTQLFFTSNDTSDYKLNQDLATATL---NLNVDD 353
 DB 101 LQESANTRTAWSGGTIMAGSRNQVETFTS--TLFDKNGVREAAVATKGRIVADS 157
 QY 354 NGVSIKSDNVQVLFEGDGFHYQYEQFANGKRENDYDCLRDHPV-VQLENGDRYLVFE 412
 DB 158 EGVMEKFGDQSTDLFQADGLFYQNY-----AENLNWFRDPHFVFNPEDEGETYALFE 209
 QY 413 ANTGTEDYQSDDOIYNWANYGGDDAFENIKSFKLLNKKKDELAGLAGALGILKLTNNQ 472
 DB 210 ANVAT--VRGEDI-----GEDEIGVPVANTVV--PKD---ANLCSASIGIARCL--- 252
 QY 473 SKPKVE-EVYSPVSTLMACDEVXXK---IGDKYILFVSRVSRGSDRELTKDNTIVG 527
 DB 253 SPDRTEWELLPLLTAFGVNDQMRPHVIFQNGLTLYLFTISHDSTYAD-GLTGS----- 306
 QY 528 DNVAMIGYVSDS-LMGKYKPLNNGSVLTASVPANWRTATSYIYAVPVAGHPDOVLTSY 586
 DB 307 ---GLYGFVSENGIFGYPEPLNGSLVLGG--PASQPTAYAHYIM-----NNGLVESE 355
 QY 587 M-----SNKDFASGEGNYATWAPSLVQINPDFT 615
 DB 356 INEIIDPKSGKVIAGG-----SLAPTVRVELOGHET 386
 RESULT 7
 ID SACC_ZYMMO STANDARD; PRT; 423 AA.
 AC 060114; Q60116; Q60487;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
 DE (Sucrose 6-fructosyl transferase).
 GN SACC OR LEVU OR SUCE2.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 10988 / ZM1;
 RX MEDLINE=93305726; PubMed=8318541;
 RA Song K.B., Joo H.K., Rhee S.-K.;
 RT "Nucleotide sequence of levansucrase gene (levU) of Zymomonas mobilis
 RT ZM1 (ATCC10988).";
 RL Biochim. Biophys. Acta 1173:320-324(1993).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-IFO 13756;
 RX MEDLINE=95218269; PubMed=7766026;
 RA Kyono K., Yanase H., Tonomura K., Kawasaki H., Sakai T.;
 RT "Cloning and characterization of Zymomonas mobilis genes encoding
 RT extracellular levansucrase and invertase.";
 RL Biogsci. Biotechnol. Biochem. 59:289-293(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL B806;
 RX Gunasekaran P., Mukundan G., Kannan R., Velmurugan S.,


```

Qy 548 NNSGVLTATSVANWRRTATYSYAVPVAGHPDQVLITTSYMSNKDFASGEGNY----ATWAP 604
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 335 NGSLVL--GNPSPQRPQTYSHCVMP-----NGLVTSFDN--VPTSDGNRYRGTEAP 384
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 605 S 605
      :
Db 385 T 385

RESULT 9
SACB_PSESG STANDARD; PRT; 415 AA.
AC OS2408;
DT 13-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
DE (Sucrose 6-fructosyl transferase).
DE LSC.
OS Pseudomonas syringae (pv. glyciniae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG4180;
RX MEDLINE=98394981; PubMed=9726857;
RA Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT levansucrase genes from the plant pathogens Pseudomonas syringae pv.
RT glyciniae and P. syringae pv. phasecollicola";
RT Appl. Environ. Microbiol. 64:3180-3187(1998).
RL
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabor-
CC between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for comm-
CC entities requires a license agreement (See http://www.isb-sib.ch/annou
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF037443; AAC36056.1; -.
CC InterPro; IPR003469; Glyco_hydro_68.
CC Pfam; PF02435; Glyco_hydro_68; 1.
CC Transferase; Glycosyltransferase.
CC SEQUENCE 415 AA; 45844 MW; 839B686ACB0610CF CRC64;
CC -----
Query Match 5.8%; Score 242.5; DB 1; Length 415;
Best Local Similarity 23.6%; Pred. No. 3.2e-06;
Matches 96; Conservative 68; Mismatches 154; Indels 89; Gaps

Qy 239 VWDSPWPQDPVTGYSVNYMGYQLVIAMMG--IPNSP--TGDNIHYLL----- 281
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 44 IWDTPMLRE-LDGTVVSVANGSWISVTLTADRIIPDDPQQYVAGRYDIDKRDWEDHGRARM 102
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 282 ---YNYKGDNDFSHRNAGSIFCTKTNVQFQWSSG-S-ATVNDGGTITLPFTSNDTSDYKL 337
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 103 CYWYRTGKD----WIFGGRVNAEGVSPTRREWAGTPVLNKGIDLDLYTCVTPE----- 154
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 338 NQQLRATATLNLAVNDGNSYKSVSDNYQVLFGDGDFHYQTYEQFANGKDRENDDYCLRDP 397
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 155 --AAIAKVRGRIVTSKGVELKDFTEVTKTLEADGKYYQT-----EAQNSTWNFRDP 204
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 398 H-VVQLENGDRYLVFANTGTEDYQSDQIYNWANYGGDAFNKISSFKLLNNKKDRELA 456
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 205 SPFIDNPDKGLYVFGCNV-----AGERTGHTVGAABLGPVPGHEHT 247
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Qy 457 GLANGALGILKLTNNQSPKVE-EVYSPLYSTLMACDEVXKK-----LGDKYYLFSVTRVS 511
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 248 GGARFQVCGICGLAVAKDLSGDEWEILLPLVTAVGVNDOTERPHVVFQDCKYYLFTIS--- 304
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Qy 512 RGSDBELTAKDNTIVGDNVAMIGTVSOSLMGKVKPLNNSGVVLTASVPANWRPATVSYVA 571
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 305 ----HKFTYADGVTPGDCV--YGVFGHLPFPYRPMNASGLV-GNPPAQ-PPQTYSHCV 356
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Qy 572 VPVAGHPDQVLITSYM-----SNKDEASGEGNATWAPSLFVLOINPD 613
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 357 MP-----NGLVTFIDSVPTSGEDYRIG-----GTEAPTVRILLEGD 393
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

RESULT 10
SACB_PSESH
ID SACB_PSESH STANDARD; PRT; 431 AA.
AC O68609;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
DE (Sucrose 6-fructosyl transferase).
GN LSC.
OS Pseudomonas syringae (pv. phaseolicola).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=319;
OX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 1321;
RX MEDLINE=98394981; PubMed=9726857;
RA Hettwer U., Jacek F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT levansucrase genes from the plant pathogens Pseudomonas syringae pv.
RT glycinea and P. syringae pv. phaseolicola.";
RL Appl. Environ. Microbiol. 64:3180-3187(1998).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N-1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL Outstat-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC entities requires a license agreement (See http://www.isb-sib.ch/anno-
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF052289; AAC36063.1; -.
CC InterPro: IPR003469; Glyco_hydro_68.
CC Pfam: PF02435; Glyco_hydro_68; 1.
CC Transference; Glycosyltransferase.
CC SW SEQUENCE 431 AA; 47603 MW; 795FEA246D40C40F CRG64;
CC SEQUENCE

```

	Query Match	5.8%	Score 240.5;	DB 1;	Length 431;
	Best Local Similarity	23.3%;	Pred. No. 4.2e-06;		
	Matches	95;	Conservative 68;	Mismatches 154;	Indels 91; Gaps
Qy	239	VWDSWPVQDPVTGYSVNMGYQLVIAMWG--IPNSPTGDNHIYLLYN-----KY 285			
	:	:::	:	:	:
	:	:	:	:	:
Db	60	IWDTPMLRE-LDGTPTVSVMNGSVLTLTADRHNDPQ-----YLDANGRYDIKRDWEDRH 113			
	:	:	:	:	:
Qy	286	GDNDFSHWRN-----AGSIFGKTETNYFQEWSSGAI-VNDDGTIQLFTTSNDTSDYK 336			
	:	:	:	:	:
Db	114	GRARMSYWSRTGDKWIFGGFVMAEGYSPTREWAGTPIILLNDKGDLIDLYTCVTPG--- 170			
	:	:	:	:	:
Qy	337	LNDQRLTATILNLNVDDNGSVISKVNYQVLFGDGFWHTYEQFANGKDRENDYCLDR 396			
	:	:	:	:	:
Db	171	---AAIAKVRGRIVTSQQGVLELKDFTOVKKLFDAGDTTYOT-----EAOSSNNFRD 219			
	:	:	:	:	:

```

QY 397 PH-VVLENGDRYLVEPANTCTEDYQSDDOIYNWANYGDDAFNIKSSFKLLNNKKDREL 455
DB 220 PPSFIDPNBGLKLYMVEGNY-----AGERSGSHTVGAELGVPVPPGHED 262
QY 456 AGLANGALGTLKLTNNOSKPVE-EVYSPLVSTLMACDEVXXK-----LGDKYLYLFSVTRV 510
DB 263 VGARQVQCGICGLAVAKDLGSEWEILPLVYTAGVNDQTERPHYVFDQDKYLYLFISHK 322
QY 511 SRGSDRELTAkontIVGDNVAMIGYVSDLSMKYKPLNNSGVVLTASVPANMRTATYSY 570
DB 323 FTYAE-GLTGPD-----GVYGFGEHLFGPYRPMNASGLVL--GNPPEQPFQYSHC 371
QY 571 ANPVAGHPDQVLITSYMSN-----KDFASGEGNATWAPSFVLQVNP 613
DB 372 VMP-----NGLVTSFIDSVPTEGEDYRIG-----GTEAPTVRILLKGD 409

RESULT 11
SACB_ERWAM STANDARD; PRT; 415 AA.
AC Q46654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levanisucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
DE (sucrose 6-fructosyl transferase).
GN LSC.
OS Erwinia amylovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
OX NCBI_TaxID=552;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EA774;
RA Geier G., Geider K.K.;
RT "Characterization and influence on virulence of the levanisucrase gene
from the fireblight pathogen Erwinia amylovora.";
RL Physiol. Mol. Plant Pathol. 42:387-404(1993).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X75079; CAAS2972.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 415 AA; 46408 MW; 4FF564F6E0607FEB CRC64;

Query Match 5.7%; Score 236.5; DB 1; Length 415;
Best Local Similarity 23.4%; Pred. No. 6.6e-06;
Matches 96; Conservative 67; Mismatches 152; Indels 95; Gaps 20;

QY 239 WDSNPVQDPVTGVSVNMGVQLVIANMGIPIPSPT-----GDHNI 278
DB 44 IWDTPMLRD-FDGEIISVNGWCIIIFTAORTNPQFQDENGNDYTRWDHRHGRARI 102
QY 279 YLLNXYKGDNDHSHWRNAGSIFGKTNVQFQWGSAL-VNDGDTIQIIFTSNDTSYKL 337
DB 103 CYWTSRTGKD---WIFGGRYMAEGVAPTTREWAGTPIILLNDRGDIIDLYTCVTPG--- 154
QY 338 NDQRLATATLNLVNDNGVSTKSDVNTYQVLFEGGDFHYQTEQFANGKDRNDYCLRDP 397

```

```

DB 155 --ATIAKVRGKIYTSQSVSLBFGFOQVTSLSADGTIYQTEBQ-----NAFWNRDP 204
QY 398 H-VVLENGDRYLVEPANTCTEDYQSDDOIYNWANYGDDAFNIKSSFKLLNNKKD 452
DB 205 SPFIDRNDGKLYMLFEGNVAGPRGSHEITQ-----AEMG-----NVPFGYEDVGAKY 252
QY 453 RELAGLANGALGTLKLTNNOSKPVEEYSPVSTLMACDEVXXK-----LGDKYLYLFSVT 508
DB 253 QA-----CGVG-LAVAKDLGSE-WQILPLPLITAVGVNDQTERPHVFDQDKYLYLFTIS 304
QY 509 RVSRGSDRELTAkontIVGDNVAMIGYVSDLSMKYKPLNNSGVVLTASVPANMRTATYS 568
DB 305 -----HKYTFADNLGTGPDGV--YGFVSDKLTGPTPMNSSLVL--GNPSPQFQYTS 353
QY 569 YVAVPVAGHPDQVLITSYMSN-----KDFASGEGNATWAPSFVLQVNP 613
DB 354 HYVMP-----NGLVTSFIDSVPWKGDYRIG-----GTEAPTVKILLKGD 393

RESULT 12
SACB_ACEDI STANDARD; PRT; 584 AA.
AC Q43998;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Levanisucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transferase) (sucrose 6-fructosyl transferase).
GN LSDA.
OS Acetobacter diazotrophicus (Glucanacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Glucanacetobacter.
OX NCBI_TaxID=33996;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 52-61.
RC STRAIN=SRT4 / CBS 550.94;
RA MEDLINE=96253999; PubMed=8704949;
RA Arieta J., Hernandez L., Coego A., Suarez V., Balmori E.,
Mendez C., Petit-Glatron M.-F., Lambert R., Selman-Housein G.;
"molecular characterization of the levanisucrase gene from the
endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4.";
Microbiology 142:1077-1085(1996).
CC -1- FUNCTION: RELEASES FRUCTOOLIGOSACCHARIDES AND LEVAN, A HIGH-
MOLECULAR-MASS FRUCTOSYL POLYMER, FROM SUCROSE. IT ACTS MORE AS A
SUCROSE HYDROLASE THAN AS A FRUCTAN POLYMERASE.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: THE N-TERMINAL IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB EMBL; L41732; AAB36606.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT PROPEP 31 51
FT CHAIN 52 584 LEVANISUCRASE.
SQ SEQUENCE 584 AA; 63525 MW; CBFBF4139AD0B08CE CRC64;

Query Match 4.7%; Score 197; DB 1; Length 584;
Best Local Similarity 23.8%; Pred. No. 0.0013;
Matches 117; Conservative 78; Mismatches 175; Indels 122; Gaps 31;

```

QY 220 NMPATYTDQTKMAHLD--VMSWPDQVPTVGYVSNMGYQLVITAMGIPNSPTG--D 275
 Db 119 NIPADFPV-----INPDVWVDWTLLDKHADQFS-YNGWEVIFCLTADPNAGYFDD 170
 QY 276 NHYI-----LYNK-----YCNDFSHRNAGSFGTKETNVFQWSSAI 316
 Db 171 RHVHARIGFYRRAGIPASRRPVNGWYGGHLPDQASAOVYAGQTYTNO-AEWSGSSR 229
 QY 317 VND-DGTIOLFT----SNDTSYKLN-DQRLATATL-NLNVDNDGVSISKSDVNYQVLF 368
 Db 230 LMQHGNVSVFYDVFANRDANANNITPQAITQTLGRHADFNHWFTGPTAHTPLL 289
 QY 369 EGDGFHYQTYEQFANGKRENDYCLDRPHVQ--LENGDRYLVFEANT-----GTEDYQS 422
 Db 290 QPDGVLYQ-----NG-AQNEEFNRPDFTFEDPKHPGVNMYVEGNTAGQRCVANCTE 341
 QY 423 DDQIYNWANYGGDDAFNIKSFLLNKKORELA--GLANGALILKLTNNQSKPKVEEV 480
 Db 342 AD-----LGRPNPD-NAETLQELVLDGAYYQKANIGLA-----IATDSTLSKWK---F 386
 QY 481 YSLVSTLMACDEYXXKL-----GDKYLFVS---TRVSRGSDRELTAKNITVGDNVAMI 533
 Db 387 LSPILSANCVDQTERQVYLHNKGYIFITISHRTTFAAGVD-----GPD-GVY 434
 QY 534 GYVSDSLMGKYKPLNNSGVVLTSVPAWRTA-----TYSYVAVPVAG 576
 Db 435 GFVCGIRSDFPQV-NYGSGLTMGNPTDLNTAAGTDEDPSPQNPRAFQSYSHYVMPGG- 492
 QY 577 HPDQVLTYSNKNDFASGEGNYATWAPSLVQINPDPTTVTLARATNOGDVWDD--SS 634
 Db 493 -----LVESFTDYENRRG-----GTLAPTVRVRI-AQNASAVDURYNGGLGYGDIPAN 542
 QY 635 RND-NMLGVLKE 645
 Db 543 RADVNIAGFIOD 554

RESULT 13
 ID KYK1_DICDI STANDARD; PRT; 1584 AA.
 AC P18160; 1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Non-receptor tyrosine kinase spore lysis A (EC 2.7.1.112) (Tyrosine-
 protein kinase 1).
 GN PKA OR SPLA OR DPYK1.
 OS Dictyostelium discoideum (Slime mold). Dictyostelium.
 OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JH10;
 RX MEDLINE=97053827; PubMed=8898241;
 RA Nuckolls G.H., Oshero N., Loomis W.F., Spudich J.A.;
 RT "The Dictyostelium dual-specificity kinase spla is essential for
 RT spore differentiation."
 RL Development 122:3295-3305(1996).
 RN [2]
 RP SEQUENCE OF 1248-1584 FROM N.A.
 RX MEDLINE=90287147; PubMed=1972546;
 RA Tan J.L., Spudich J.A.;
 RT "Developmentally regulated protein-tyrosine kinase genes in
 RT Dictyostelium discoideum."
 RL Mol. Cell. Biol. 10:3578-3583(1990).
 CC -1- FUNCTION: ESSENTIAL FOR SPORE DIFFERENTIATION.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT WITH A PEAK
 CC DURING THE MOUND STAGE OF MORPHOGENESIS.
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES BUT ALSO TO
 CC SERINE/THREONINE PROTEIN KINASES.

CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32174; AAB41125.1; --
 CC EMBL: M33785; AAA33202.1; --
 CC PIR: A35670; A35670.
 CC HSP: P11362; IAGW.
 CC DictyDb; DD03010; PYKA.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR001660; SAM.
 CC InterPro; IPR003877; SPRY.
 CC InterPro; IPR003878; SPRY_domain.
 CC InterPro; IPR004040; STY_pkinase.
 CC InterPro; IPR001245; Tyr_pkinase.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00536; SAM; 1.
 CC Pfam; PF00622; SPRY; 3.
 CC SMART; SM00454; SAM; 1.
 CC SMART; SM00449; SPRY; 3.
 CC SMART; SM00221; STYK; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE; PS00105; SAM_DOMAIN; 1.
 CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 KW DOMAIN 908 972 SAM.
 FT DOMAIN 403 420 POLY-ASN.
 FT DOMAIN 428 435 POLY-THR.
 FT DOMAIN 449 480 POLY-ASN.
 FT DOMAIN 483 491 POLY-ASN.
 FT DOMAIN 494 508 POLY-ASN.
 FT DOMAIN 512 532 POLY-ASN.
 FT DOMAIN 596 600 POLY-ASN.
 FT DOMAIN 808 811 POLY-PHE.
 FT DOMAIN 1026 1029 POLY-SER.
 FT DOMAIN 1195 1210 POLY-ASN.
 FT DOMAIN 1215 1220 POLY-GLN.
 FT DOMAIN 1224 1233 POLY-GLN.
 FT DOMAIN 1266 1274 POLY-PRO.
 FT DOMAIN 1289 1561 PROTEIN_KINASE.
 FT NP_BIND 1295 1303 ATP (BY SIMILARITY).
 FT BINDING 1316 1316 ATP (BY SIMILARITY).
 FT ACT_SITE 1417 1417 BY SIMILARITY.
 FT CONFLICT 1248 1248 D -> R (IN REF. 2).
 FT CONFLICT 1435 1435 V -> L (IN REF. 2).
 SQ SEQUENCE 1584 AA; 174304 MW; 5D1589458D8E01E3 CRC64;

Query Match 4.6%; Score 190.5; DB 1; Length 1584;
 Best Local Similarity 19.3%; Pred. No. 0.011;
 Matches 165; Conservative 103; Mismatches 283; Indels 303; Gaps 40;
 QY 15 SASILMGVVTAHADQVESNNYNGVAEVTQANQIGVDGKIISANSNTTSGST---N 71
 Db 396 STSIL-----VPGNNNNNNNNNNNNNNIIG-NGKITTITTTSTSPSSINN 443
 QY 72 QESSATNTENAVVNESKNTNTENAVVNEKNNTNTENAVVNEKNNTNTENDNSQLK 131
 Db 444 EDITSSNNNNNN--NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 501
 QY 132 TNEQPSAQTQANLKLNPQAAKAVQ-----AKIDAGSLTDQINEL--NKINFSK 181
 Db 502 NNN 561
 QY 182 SAEKGAKLTKDLEGICN--AIVKQ-DPOVAIPYNAKEIKKNMPATYTVDAQTGRMAHLD 238

RT (dextranase) gene results in altered adherence and dextran
 RL catabolism.", 141:2929-2936(1995).
 CC -1- FUNCTION: MAY PLAY A ROLE IN SUCROSE-INDEPENDENT ADHERENCE TO THE
 CC PELLICLE-COATED TOOTH SURFACE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,6-alpha-D-glucosidic
 CC linkages in dextran.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 CC IN THE REGION OF THE MEMBRANE ANCHOR.
 CC -1- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D49430; BAA08409.1; .
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
 KW Hydrolase; Glycosidase; Signal; Transmembrane.
 FT SIGNAL 1 ?
 FT CHAIN ? 850 DEXTRANASE.
 FT DOMAIN ? 823 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 824 844 POTENTIAL.
 FT DOMAIN 845 850 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 811 816 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 SQ SEQUENCE 850 AA; 94538 MW; D653CAA39C93D2AC CRC64;

Query Match 4.5%; Score 185; DB 1; Length 850;
 Best Local Similarity 21.0%; Pred. No. 0.0094;
 Matches 201; Conservative 108; Mismatches 290; Indels 358; Gaps 55;

Qy 26 AHADQVESNNTNGVAEVNTERQANGQIGVDGKIISANSNT-----TSGSTNOESS---A 76
 Db 40 AAAYQTES-----ASIDSNEQEQS-----VSANTNTPQAKKLSNNSHQEPQMVMYS 85
 Qy 77 TNNTENAVVNESKNTNN-----TENAV-----VNEKNKNTNTENAVVNENKNTN--N 121
 Db 86 AANKERAVLETAQKQKNKNMNLTTDKAVYQAGEAVHLNLTNNTSLAQNATAVEVYS 145
 Qy 122 TENDNSQLKLTNNEQPSAATOANLKKLPQAKAVQNAK--IDAGSLTDDQINELNKNIF 179
 Db 146 LENLKLTLQYT-----KYLPPNESYTTQKGEFVIPANSLANNR-GYLLKVN 191
 Qy 180 SKS-----AEKGAKL-----TFKDLGIG-----NAIVKQDPQYAIYSNAKEIKN 220
 Db 192 SDSQNNILEQGNRAIATEDDWRTEFPYAAIGSGKDNNSVLTKNLPDY---YRELEQMKN 248
 Qy 221 MPAT-----YTV-----DA-----QTGKMAHLDV 239
 Db 249 MNINSYFYDVYKSNTPFPNVPKFDQSNWNWSHVSQVETDAVKALVNRVHQGVAML-- 306
 Qy 240 WDSWFPVQD-----PVTGYVSNTM--GYQLVIAMGIPNSPTGDNHIYLLYN-----K 284
 Db 307 YNMILAQANETAVLPDTEIYNYETGY-----GQNGQVMYTSIDDKPLQ 352
 Qy 285 YGDNDFS-HWRN--AGSIFGTKETVWFQWEGSAIVNDDGTQLPFTSNDTSD-----YKL 337
 Db 353 YYNPLSKSWONYISNAMAQAMKTKGFCWQGDITI--GDNRY-LSHNQKDSRDIAHSEFML 409
 Qy 338 ND-----ORLATATLNLVNDNGVSIKSDVN-----YQVL--PEGDGFHYQTYE 379
 Db 410 SDVYAEFLNKKMKELPQYLLTN-DVNGENISKLSKQDVIYNELWPGTSALGNRPOE 468
 Qy 380 QFANGKDRENDDYCLRPVQVLENGDRYLV-----FEAN-----TCTEDY 420

Search completed: September 26, 2002, 18:26:47
 Job time: 186 sec

Db 469 SYGDLKARVDQ-----VRQATGKSLIVGAYMEEPKFDDNRPIPLNGAARDVLASATY 519
 Qy 421 QSDQDYNNWANYGGDDAFNFKSFKLLNNKKORELAGLAN-----GALGILKLTNNOSKPK 476
 Db 520 QTDVAVLLTTAAIAAAGGYHM-----SLAALANPNDGGGVGL----- 556
 Qy 477 VEEVYSPLVSTLMACDEVXXKLGDKYVLFVSVTRSVSRGSDRELAKDNTIVGDNV----- 530
 Db 557 -ETAYPTQSL-----KVSKELNKNYHY-----QQFITAYEN-LURDKVENDSAE 600
 Qy 531 -----AMIGVYSDSLMGKYKPLNNSGVVLTASVPANMRTATYSYAYVPVAGHPDQVL--- 582
 Db 601 PQTFETANGRLSODALG-----INGDQVWTYAKKGNDFRTI-----QLLNLM 642
 Qy 583 -TTSYMSNKDFASGEGNYATWAPSLVQINPDDTTTVLAR-----A 622
 Db 643 GITSWKNEDE--GYENNKNT-----PDEQTNLLVTVPLTGVSMAEADRIAKQVYL 689
 Qy 623 TNOGDVWDDSSRRNDNMGLVYLKEGAANSALPGEWGKPV-----DWSLINRSPG 671
 Db 690 TSPDDWL-----QSSMISLTTQVKTN-----ENGDPVLYIQVPRITLMDMIYINET-- 735
 Qy 672 LGLPHQPVQPKIDQDQDPQSPGONTKNVTPGNGDKPKAGKATPDNTNIDPSAQSPSQNTNI 731
 Db 736 --IKPETPKVP--EQPOH-----PARTLEPAIPOTPEAVSLPVANKQAEQGNK 780
 Qy 732 DPSAQX-SQONTKNVTPGNEKQKNTDAKOLPOTGNKSGIAGLYAGSLALFLGLAAI 787
 Db 781 NELVSALTGEENDLQLPTLSKRSLSISQAEPLQTDGNNET-----RSNLLKVIAGAGAL 833

Db 403 PHVIEDNGORYLVEASTGTENTYQEDQIYNFTNYGSSAYNVKSLFRFLDODMYNRA 462
QY 457 GLANGALGILKLTNNOSKPKVEEYVSLVTLMACDEV-----XXKLGDKYLFVSVTRVSR 512
Db 463 SWANAALGILKLGDKKTPEDQFYTPLLSSVSDSELRPNVYKLGDKYLFVTSRLNH 522
QY 513 GSORELTAQDNTIVGDNVAMIGVSVDSLMGKYYKPLNNSGVVLTASVPANRRTATYSYAV 572
Db 523 GSNDAHNAKANEVGDNVMLGVSDQLTNGYKPLNNSGVVLTASVPADNRTATYSYAV 582
QY 573 PVAGHPDOVLITSYMSKNKDFASCEGNYATWAPSLFVQINDDTTVLARATNGDWMYDD 632
Db 583 PVAGSSDTLLMTAYMTRNEVAGKGNSTWAPSLFQVLDPGTTKVALEMTQQGDWIDE 642
QY 633 SSRNDMLGVLEKGAANSALPEWCKPVDWSLINSPLGLKPHQVOPKIDPOQOPS 692
Db 643 PSRTTDTVGTL-----DTAVLPENGDYIDNVIG---GYGLAPHTPGQ-----YQPT 687
QY 693 GQNT-----KNVTPGNGDKPAGKATPDNTNIDPSAOPSGQ 727
Db 688 VPSTPIHTDDIISFEVSFDGHLVLPKPVNNDNSAGRIDQSRNSGGSINLVAFNVA---GG 744
QY 728 NTNIDPSAOSGONTKNVTPGNEKCKNTDAKOLPOTGNKSGLAGLYAGSLALF 782
Db 745 NISVKFS-QKSNINTKET-----KKAHVSTKQKKGNS-----FFAALLALF 787
RESULT 2
JQ0802
levansucrase (EC 2.4.1.10) precursor - Bacillus amyloliquefaciens
N:Alternate names: 2,6-beta-D-fructan-6-beta-D-fructosyltransferase; sucrose
C:Species: Bacillus amyloliquefaciens
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 15-Oct-1999
C:Accession: JQ0802; S11739
R:Tang, L.B.; Lenstra, R.; Borchert, T.V.; Nagatajan, V.
Gene 96, 89-93, 1990
A:Title: Isolation and characterization of levansucrase-encoding gene from Bacillus amyl
A:Reference number: JQ0802; MUID:91092506
A:Accession: JQ0802
A:Molecule type: DNA
A:Residues: 1-472 <N>
A:Cross-references: EMBL:X52988; NID:g39333; PIDN:CAA3179.1; PID:g39334
A:Experimental source: ATCC 23844
C:Genetics:
A:Keywords: glycosyltransferase; hexosyltransferase
F:1-29/domain: signal sequence #status predicted <SIG>
F:30-472/Product: levansucrase #status predicted <NAT>
Query Match 18.1%; Score 751.5; DB 2; Length 472;
Best Local Similarity 38.0%; Pred. No. 2.2e-34;
Matches 190; Conservative 88; Mismatches 171; Indels 51; Gaps 19;
QY 144 NLKLNQAKAVONAKIDAGSLTDQINELNKNIFSKSAEKGAKLTKDLGEGICNAIVK 203
Db 2 NIKKIVKQATVLTETALLAGATQAFAKENKQAY-KETGVSVHITHMDLIQIPKQ--Q 58
QY 204 QDPQVAPISNAKEKNPAPYTVDAQTGKMAHLVDWSPVQDPVTGVSYNMGYQLVI 263
Db 59 QNEKYQVPQFDOSTIKN-----IESAKG---LDWDSWPLQN-ADGTVAEYNGHYVVF 107
QY 264 AMMGIPNSPTGDHHTLYLKYNGDQFHSWRNAGSIFGTETNVP-----QEMS 312
Db 108 ALAGSPKD-ADTSLYMFYQKVGNSIDSWKNAGRVF--KDSKDFDANDPILKDOTQEWS 164
QY 313 GSAIYNDGDTIQLEFSTNDTSDYKLNQDLQATLNLNVDDNGSVISKVDNVOVLFEGDG 372
Db 165 GSATFTSGKIRLFYV--DYSGKHGKQSLTAQVNSKSDDTLKINGVEDHKTIFDGDG 222
QY 373 PHYQTYEAFN-GKORENDYCYLRDPHVVQLENGDRYLVFEANTGTED-YQSDQIYNWA 430
Db 223 KTYQNVVOOFIDEGNVTSGDNHTRDPHYVE-DKGHKYLVFEANTGTENGYGESLFPKA 281

QY 431 NYGDDAFNIKSPKLLNNKKDRELAGLANGALGILKLTNNOSKPKVEEYVSLVSTLMA 490-
Db 282 YYGGTNFFRKESQKLOOSAKRD-AELANGALGILIELNNDYT---LKKVKKPLITSNTV 337
QY 491 CDEV-----XXKLGDKYLFVSVTRVSRGSDRELTAQDNTIVGDNVAMIGVSVDSLMGKYPK 546
Db 338 TDETERANVFKNMGKWLFT---DSRGSKMTI-----DGINSNDIYMLGYVNSLTGPYKP 390
QY 547 LNSGVVLTASVPANRRTATYSYAVPVAGHPDOVLITSYMSKNKDFASGEGNTATWAPSF 606
Db 391 LNKTLGLVQLMGDPNDVFTTSHFAVPOA-KGNNVITSYMTNRRGFF--EDKKATGQPSF 447
QY 607 LVQINPDQTTTLARATNGQ 626
Db 448 LMKIKGNKTSVVKNSILEQG 467
RESULT 3
A25040
levansucrase (EC 2.4.1.10) sacB precursor - Bacillus subtilis
N:Alternate names: sucrose 6-fructosyl-transferase
C:Species: Bacillus subtilis
C:Date: 16-Aug-1988 #sequence_revision 20-Feb-1995 #text_change 20-Jun-2000
C:Accession: S07309; A25040; I39967; H69702
R:Steinmetz, M.; le Coq, D.; Aymerich, S.; Gonzy-Treboul, G.; Gay, P.
Mol. Gen. Genet. 200, 220-228, 1985
A:Title: The DNA sequence of the gene for the secreted Bacillus subtilis enzyme levan
A:Reference number: S07309; MUID:85295507
A:Accession: S07309
A:Molecule type: DNA
A:Residues: 1-473 <STE>
A:Cross-references: EMBL:X02730; NID:g40118; PIDN:CRA26513.1; PID:g732568
A:Experimental source: Marburg
R:Shimotsu, H.; Henner, D.J.
J. Bacteriol. 168, 380-388, 1986
A:Title: Modulation of Bacillus subtilis levansucrase gene expression by sucrose and
A:Reference number: A25040; MUID:87008406
A:Accession: A25040
A:Molecule type: DNA
A:Residues: 1-68 <SHI>
A:Cross-references: GI:M14202; NID:g143485; PIDN:AAA22725.1; PID:g143486
R:Fouet, A.; Arnaud, M.; Klier, A.; Rapoport, G.
Biochem. Biophys. Res. Commun. 119, 795-800, 1984
A:Title: Characterization of the precursor form of the exocellular levansucrase from B
A:Reference number: I39967; MUID:84178454
A:Accession: I39967
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-11,1',13-62 <RES>
A:Cross-references: GB:K01987; NID:g143483; PIDN:AAA22724.1; PID:g143484
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gal
leeh, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y., M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schlicht, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: H69702
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-473 <KUN>
A:Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15450.1; PID:g26359
A:Experimental source: strain 168

C:Genetics:

A:Gene: sacB
A:Keywords: extracellular protein; glycosyltransferase; hexosyltransferase
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-473/Product: levansucrase #status predicted <MAT>

Query Match 18.0%; Score 746.5; DB 2; Length 473;
Best Local Similarity 36.3%; Pred. No. 4.2e-34;
Matches 185; Conservative 92; Mismatches 162; Indels 71; Gaps 18;

QY 144 NLKLNPOAKAVONAKIDAGSITDDQINELNKFNSKSAEKGAKLTFKDLLEGIGNAIVK 203
DB 2 NIKKFAKQATVLTFTTALLAGGATQAFKATN-----QPKYETGISH-ITR 48
QY 204 QD-----POYAIPYNAKEIKNMPATYTVDAQTGKMAHLVDWSPVQDPVTGYV 253
DB 49 HMLQIPEQKNEKYQVPEFSDSTIKNISSA-----KGLDVWDSWPLQN-ADGTV 97
QY 254 SNYMGYQVLVIAMGTPNSTGDNHLYLLNKGNDNDFSHRNAGSIFGKTETNVF----- 308
DB 98 ANYHGYHIIFALAGDPKN-ADDTSIYMFYQKVGETSIDSWKNAGRVFE--KDSKDFDANDS 154
QY 309 -----QEMSGSAIVNDGCTIOLFFTSNDTSYKLNQRLATATNLNVDNDGVSISKVD 362
DB 155 ILKDOTQEMSGSATFTSDGKIRLFYF--DFSGKHGKQTLTTAQVNVASDSSLNNGVE 212
QY 363 NYQVLFEGDGFHYQYEQFAN-GKRENDYDCLRDHPVHVOLENGDRYLVEFANTGTED-Y 420
DB 213 DYKSFIDGDKTYQNVQFIDEGNYSNGDNHTLRDPHYVE-DGKHLYLVEFANTGTEDGY 271
QY 421 QSDQIYNWANGDDAFNIKSFLLNKKRELALAGLILKLTNNQSKPKVEEV 480
DB 272 QGEESLFNKAYYKSTSFRRQESKILQSDKKR-TAELANGALMIELNDDYT---LKKY 327
QY 481 YSPLVSTLMACDEV---XXKLGDKYLFVSFVRSRGSDELTAQNTIVGDNVAMITGVV 536
DB 328 MKPLIASNTVTDEIERANVFKNKGYLFTDSRGSKMTIDGITSND-----IYMLGYV 380
QY 537 SDSLMGKYPLNNSGYLVFASVPANWRTATYSYAVPVAGHPDQVLTYSYMSKDFASGE 596
DB 381 SNSLTGPKYPLNKLGLVKMDLPNDVDTFTYSHFAVPAQ-KGNVVTYSYMTNRGEVADK 439
QY 597 GNYATWAPSLVQINPDDTTVLARATNQG 626
DB 440 --OSTFAPSLNKGKTSVVKDSILEQG 467
RESULT 4
F97118
levansucrase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97118
R: Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97118
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79737.1; PID:g15024742; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1772

Query Match 16.5%; Score 685.5; DB 2; Length 489;
Best Local Similarity 33.5%; Pred. No. 1.1e-30;
Matches 170; Conservative 101; Mismatches 167; Indels 69; Gaps 18;

QY 161 IDAGSLTDDQINELNKFNSKSAEKGAKLTFKDLLEGIGN-----AIVKQDPQYAI 211
DB 16 ICSGIITQOTFASTNDKN-----YKETYGVSHITRYNMKIPMEQNDLKEKVP 63
QY 212 YSNAKEIKNMPATYTVDAQTGKMAHLVDWSPVQDPVTGYVSNYMGYQVLVIAMGIPNS 271
DB 64 QFNASTLKNNTASAKGYD-KGNLIDLDVWDSWPLONG-DGTVANYHGYHIIFALAGDPKN 121
QY 272 PTGDNHLYLLNKGNDNDFSHRNAGSIFGKTETNV-----FOEMSGSAIVNDGCT 322
DB 122 -QDDTSIYMFYQKVGETSIDSWKNAGRVFE--KDSKDFDANDS 180
QY 323 IOLFFT-----SNDTSYKLNQRLATATNLN-VDDNGVSIKSDVNYQVLEF-GD 371
DB 181 VRLFYTFDSGVAKDGGTDAS-----NOVITTVQNLNVDNDGVSISKVD 235
QY 372 GFHYQYEQFAN-GKRENDYDCLRDHPVHVOLENGDRYLVEFANTGTED-YQSDQIYNW 429
DB 236 GTIYQNVQFIDEGKWSGDNHTLRDPHYVE-DGKHLYLVEFANTGTEDGYOGLTSLNK 294
QY 430 ANYGDDAFNIKSFLLNKKRELALAGLILKLTNNQSKPKVEEVYSPVSTLM 489
DB 295 AFYGRSOSFFKTEKQLLIDTNKKHDSANGLALITELNDDYT---LKKEMKPLIASNT 351
QY 490 ACDEV---XXKLGDKYLFVSFVRSRGSDELTAQNTIVGDNVAMITGVVSDSLMGKYK 545
DB 352 VTDEIERANVFKNKGYLFTDSRGSKMTIDGITSND-----IYMLGFSNLSLTGPKY 404
QY 546 PLNNSGYLVFASVPANWRTATYSYAVPVAGHPDQVLTYSYMSKDFASGEVATWAPS 605
DB 405 PLNKTGLVLNLDPTDLTFTYSHFAVPAQ-KGNVVTYSYMTNRGMYE--DHSSSPAPS 461
QY 606 FLVQINPDDTTVLARATNQGDNWDD 632
DB 462 FLNKGKTSVVKDSILEQG 488
RESULT 5
H97118
levansucrase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97118
R: Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97118
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79737.1; PID:g15024744; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1774

Query Match 7.6%; Score 315; DB 2; Length 428;
Best Local Similarity 27.1%; Pred. No. 3.1e-10;
Matches 105; Conservative 68; Mismatches 158; Indels 56; Gaps 18;

QY 236 HLDVWDSWSPVQDPVTGYVSNYMGYQVLVIAMGIPNSPTGDNHLYLLNKGNDNDFSHRN 295
DB 66 NLWVMDTWPVKK-DGSLAVVNGYKVIIFALTASRVGNKRDHVDAGISYFCSTGDENWY 124
QY 296 AGSLFGTKETNVFOEMSGSAIVNDGCTIOLFFTSNDTSYKLN--DORLATATNLNVD 353
DB 125 KGLAYNVEDALGSRQWAGSAILDENGWVQFFYATGRKEAVRTFEQRLVTKTFSLNVDK 184
QY 354 NGVSTSYVDNTQVLEPGDGFHYQYEQFANGKRENDYDCLRDHPVHVOLENGDR-YLVEF 412
DB 185 GGVHITNCKSHQVILEPVGVIYQTMQO-AKGP----IYSFDPYFFEDPKTKOYLIFE 239

QY 413 ANTCTE-DYQSDQIYNWANYGGDAFNKSSFKLLNNKDKRELALANGALGIL----- 466
DB 240 GNKGKIEKMPENI-----GDKLPR-----KNHAPRGVENF-NGNNGVIAQNK 284
QY 467 KLTNNQSPKPYEYVPLVSTLMACDEYXKLGKYYLFSVTR-----VSRGSDRELTKADN 523
DB 285 DLTRFKLLPPLLEAVG--VNOOLRPPQIVMK-KNKYYLFTISHKFTYAQG-----LNGVD- 336
QY 524 TIVGDNVAMIGYVSDSLMGKYLNNNGSVLTASVPANNRRTATSYAVPVAGHPDQVLI 583
DB 337 -----GLYFCGNSLRNKPPLNGLVITN--PTNDPQTYSWYL--VSGHDVLSFI 385
QY 584 TSYMSNKDFASGEGNYATWAPSFVLQI 610
DB 386 NEYHFNGLRYG----GTFAPTQLQISL 408

RESULT 6

JC2520
beta-fructofuranosidase (EC 3.2.1.26) - Zymomonas mobilis
N:Alternate names: Invertase
C:Species: Zymomonas mobilis
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Jun-2000
C:Accession: JC2520; PC2377; S50037
R:Kyono, K.; Yanase, H.; Tomomura, K.; Kawasaki, H.; Sakai, T.
A:Title: Cloning and characterization of Zymomonas mobilis genes encoding extracellular
A:Reference number: JC2519; MUID:95218269
A:Accession: JC2520
A:Molecule type: DNA
A:Residues: 1-413 <KYO>
A:Cross-references: DDBJ:D17524; NID:g809529; PIDN:BAA04476.1; PID:g809531
A:Accession: PC2377
A:Molecule type: protein
A:Residues: 1-21 <KY2>
R:Rhee, S.K.
Submitted to the EMBL Data Library, March 1994
A:Reference number: S50037
A:Accession: S50037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-104, 'S', 106-107, 'T', 109-413 <RHE>
A:Cross-references: EMBL:L08094; NID:g468429; PIDN:AAA61488.1; PID:g468430
C:Genetics:
A:Gene: sucZE3
C:Keywords: glycosidase; hydrolase

Query Match 7.48; Score 308; DB 2; Length 413;
Best Local Similarity 26.5%; Pred. No. 7.3e-10;
Matches 105; Conservative 72; Mismatches 149; Indels 70; Gaps 20;

QY 239 VWSMPVDPVTGYVSNYMGYQVLVIAMGIPNSPTGDNHLYLLNKKYDNDFSHWRNAGS 298
DB 42 LMDTWPLRD-INGNPVSEKGNVIFSLVADRNPWDRSHARIGYFYSKDGKSWVYGGH 100
QY 299 IFGKTETNVFOEWSGSAIV--NDDGTIQLFFTSNDTSYKLNDRLATATL---NLNVDD 353
DB 101 LLQELANSRTAEWSGGTIMAPGSRNOVETFTS---TLFDKNGVREAAVATKGRYADS 157
QY 354 NGVSIKSDVNYQVLFEGDGFHYQTEQFANGKDRNDYCLDRPHV-VOLENGDRYLVE 412
DB 158 EGVWFKGFQDSTDLFQADGLFYQNY-----AENLNFRDPHFVFINPEDEGTALFE 209
QY 413 ANTGTEDYQSDQIYNWANYGGDAFNKSSFKLLNNKDKRELALANGALGILKLTNNQ 472
DB 210 ANVAT--VRGEDI-----GEDEIGVPANTVV---PKD---ANLCSASIGIARCL--- 252
QY 473 SKPKVE-EVYSPVSTLMACDEVXXK-----LGDKYLFVSFVTRVSRGSDRELTKADNTIVG 527
DB 253 SPDRTEWELLPLLTAFGVNDQMERPHVIFQNGLTLYLTISHDSTYAD-GLTGSD----- 306

QY 528 DNAMIGYVSDS-LMGKYKPLNNSGVLTASVPANNRRTATSYAVPVAGHPDQVLTISY 586
DB 307 ---GLYGFVSENGIFGYPPLNGSLGLVGG--PASQTEAYAHYIM-----NNGLVESF 355
QY 587 M-----SNKDFASGEGNYATWAPSFVLQINPDDT 615
DB 356 INEIIDPKSGKVIAGG-----SLAPTVRVELOGHET 386
RESULT 7
S47527
extracellular sucrose - Zymomonas mobilis
C:Species: Zymomonas mobilis
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-1999
C:Accession: S47527
R:Song, K.B.; Lee, S.K.; Joo, H.K.; Rhee, S.K.
A:Title: Nucleotide and derived amino acid sequences of an extracellular sucrose gene
A:Reference number: S47527; MUID:94368848
A:Accession: S47527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <SON>
A:Cross-references: EMBL:L08094

Query Match 7.0%; Score 292; DB 2; Length 413;

Best Local Similarity 26.3%; Pred. No. 5.6e-09;
Matches 104; Conservative 69; Mismatches 153; Indels 70; Gaps 20;
QY 239 VWSMPVDPVTGYVSNYMGYQVLVIAMGIPNSPTGDNHLYLLNKKYDNDFSHWRNAGS 298
DB 42 LMDTWPLRD-INGNPVSEKGNVIFSLVADRNPWDRSHARIGYFYSKDGKSWVYGGH 100
QY 299 IFGKTETNVFOEWSGSAIV--NDDGTIQLFFTSNDTSYKLNDRLATATL---NLNVDD 353
DB 101 LLQELANSRTAEWSGGTIMAPGSRNOVETFTS---TLFDKNGVREAAVATKGRYADS 157
QY 354 NGVSIKSDVNYQVLFEGDGFHYQTEQFANGKDRNDYCLDRPHV-VOLENGDRYLVE 412
DB 158 EGVWFKGFQDSTDLFQADGLFYQNY-----AENLNFRDPHFVFINPEDEGTALFE 209
QY 413 ANTGTEDYQSDQIYNWANYGGDAFNKSSFKLLNNKDKRELALANGALGILKLTNNQ 472
DB 210 ANVAT--VRGEDI-----GEDEIGVPANTVV---PKD---ANLCSASIGIARCL--- 252
QY 473 SKPKVE-EVYSPVSTLMACDEVXXK-----LGDKYLFVSFVTRVSRGSDRELTKADNTIVG 527
DB 253 SPDRTEWELLPLLTAFGVNDQMERPHVIFQNGLTLYLTISHDSTYAD-GLTGSD----- 306
QY 528 DNAMIGYVSDS-LMGKYKPLNNSGVLTASVPANNRRTATSYAVPVAGHPDQVLTISY 586
DB 307 ---GLYGFVSENGIFGYPPLNGSLGLVGG--PASQTEAYAHYIM-----NNGLVESF 355
QY 587 M-----SNKDFASGEGNYATWAPSFVLQINPDDT 615
DB 356 INEIIDPKSGKVIAGG-----SLAPTVRVELOGHET 386

RESULT 8

JC2519
levansucrase (EC 2.4.1.10) precursor - Zymomonas mobilis
N:Alternate names: sucrose 6-fructosyltransferase
C:Species: Zymomonas mobilis
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Jun-2000
C:Accession: JC2519; PC2376
R:Kyono, K.; Yanase, H.; Tomomura, K.; Kawasaki, H.; Sakai, T.
A:Title: Cloning and characterization of Zymomonas mobilis genes encoding extracellular
A:Reference number: JC2519; MUID:95218269
A:Accession: JC2519
A:Molecule type: DNA
A:Residues: 1-423 <KYO>

A:Cross-references: DDBJ:D17524; NID:g809529; PIDN:BAA04475.1; PID:g809530
A:Accession: PC2376
A:Molecule type: Protein
A:Residues: 1-19 <KY2>

C:Genetics:
A:Gene: suc2E2
C:Keywords: glycosyltransferase; hexosyltransferase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-423/Product: levansucrase #status predicted <MAT>

Query Match 6.0%; Score 249.5; DB 2; Length 423;
Best Local Similarity 23.8%; Pred. No. 1.3e-06;
Matches 120; Conservative 67; Mismatches 150; Indels 167; Gaps 27;

```
QY 228 DAQTKMAHLD-----VWDSWPQDPVTGYVSNYGYQLVIAMM-----GIPN 270
DB 24 DDPATMPTIDYDFIMTDKYYWDTWPLRD--INGOVVSFGQWVSIFALVADRTKYGWHN 82
QY 271 SPTGDNHLYLLNKYKNDGDFSHRNAGSIFGTKEINVFQWSSAIV--NDGTTIQLPFT 328
DB 83 RNDG-ARIGYFYSRGSN-----WIFGGHLLKDGANPRSWESGCTIMAGTANSVEVEFT 137
QY 329 S-NDTSDYKLNDRORLATATNLNVDDNGVSIKSVNYQVLFEGDGFHYQTYEQFANGKDR 387
DB 138 SYNDTP-----SESVAQCKGYIYADKSVWFDGDKVTDLFQADGLYIADY-----A 185
QY 388 ENDDYCLDRPHV-VOLENGDRYLVEAN-----GTEDY-----QSDQIYNW 429
DB 186 ENNFDFRDPHFVFNPEDKYIALFEGNVAGVAGEEIGVPPKTPETDGAARYA 245
QY 430 ANYGDDAFN-IKSFKLL-----NNKDRLAGLANGALGILKLTNNQSKPKVEE 479
DB 246 AAIGIAQALNEARTEWKLPLPLVAFGYNDQTERPHVVFQNG-----LT----- 289
QY 480 VYSLVSTLMACDEVYKXKLGDKYKLFSTRVSRGSDRELTAKDNTIVGDNVAMIGYVSDS 539
DB 290 -----LYFTISHHSTYAD-GLSGPD-----GYGFVSEN 317
QY 540 -LMGKYKPLNNSGVVLTASVPANRMTATSYIYAVPVAGHPDQVLITSYMSNKNDFASGSGN 598
DB 318 GIGFYEPPLNGSLVL--GNPSSQPYQAYSHYV-----TNGLYTSFIDT--IPSSDPN 367
QY 599 Y-----ATWAP-----SFLVQINPDDTTVTLARATNQGDVWDDSSRNDNMLGVK 644
DB 368 VRYGGTLAPTVKLELVGHRFSVTEVKG-----YGYIPPOIWEWLAEDSSN----- 413
QY 645 EGAANSAALPGEWKPKVDWSLINR 668
DB 414 ----SAAAL-----SLLNK 423
```

RESULT 9
S33771
levansucrase - Zymomonas mobilis (ATCC 10988)
C:Species: Zymomonas mobilis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000
C:Accession: S33771

R:Song, K.B.; Joo, H.K.; Rhee, S.K.
Biochim. Biophys. Acta 1173, 320-324, 1993
A:Title: Nucleotide sequence of levansucrase gene (levu) of Zymomonas mobilis 2M1 (ATCC)
A:Reference number: S33771; MUID:93305726

A:Accession: S33771
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-423 <SON>
A:Cross-references: GB:I08093; NID:g5880348; PIDN:AAA27695.1; PID:g295699

Query Match 5.8%; Score 242.5; DB 2; Length 423;
Best Local Similarity 24.1%; Pred. No. 3.2e-06;
Matches 115; Conservative 76; Mismatches 173; Indels 113; Gaps 25;

```
QY 228 DAQTKMAHLD-----VWDSWPQDPVTGYVSNYGYQLVIAMM-----GIPN 270
DB 24 DDPATMPTIDYDFIMTDKYYWDTWPLRD--INGOVVSFGQWVSIFALVADRTKYGWHN 82
QY 271 SPTGDNHLYLLNKYKNDGDFSHRNAGSIFGTKEINVFQWSSAIV--NDGTTIQLPFT 328
DB 83 RNDG-ARIGYFYSRGSN-----WIFGGHLLKDGANPRSWESGCTIMAGTANSVEVEFT 137
QY 329 S-NDTSDYKLNDRORLATATNLNVDDNGVSIKSVNYQVLFEGDGFHYQTYEQFANGKDR 387
DB 138 SYNDTP-----SESVAQCKGYIYADKSVWFDGDKVTDLFQADGLYIADY-----A 185
QY 388 ENDDYCLDRPHV-VOLENGDRYLVEAN-----GTEDY-----QSDQIYNW 429
DB 186 ENNFDFRDPHFVFNPEDKYIALFEGNVAGVAGEEIGVPPKTPETDGAARYA 245
QY 430 ANYGDDAFN-IKSFKLL-----NNKDRLAGLANGALGILKLTNNQSKPKVEE 479
DB 246 AAIGIAQALNEARTEWKLPLPLVAFGYNDQTERPHVVFQNG-----LT----- 289
QY 480 VYSLVSTLMACDEVYKXKLGDKYKLFSTRVSRGSDRELTAKDNTIVGDNVAMIGYVSDS 539
DB 290 -----LYFTISHHSTYAD-GLSGPD-----GYGFVSEN 317
QY 540 -LMGKYKPLNNSGVVLTASVPANRMTATSYIYAVPVAGHPDQVLITSYMSNKNDFASGSGN 598
DB 318 GIGFYEPPLNGSLVL--GNPSSQPYQAYSHYV-----TNGLYTSFIDT--IPSSDPN 367
QY 599 Y-----ATWAP-----SFLVQINPDDTTVTLARATNQGDVWDDSSRNDNMLGVK 644
DB 368 VRYGGTLAPTVKLELVGHRFSVTEVKG-----YGYIPPOIWEWLAEDSSN----- 413
QY 645 EGAANSAALPGEWKPKVDWSLINR 668
DB 414 ----SAAAL-----SLLNK 423
```

RESULT 10
S39195
levansucrase - Erwinia amylovora
C:Species: Erwinia amylovora
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: S39195
R:Geier, G.; Geider, K.K.
submitted to the EMBL Data Library, September 1993
A:Description: Characterization and influence on virulence of the levansucrase gene
A:Accession: S39195
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-415 <GEI>
A:Cross-references: EMBL:X75079; NID:g433558; PIDN:CAA52972.1; PID:g433559

Query Match 5.7%; Score 236.5; DB 2; Length 415;
Best Local Similarity 23.4%; Pred. No. 6.8e-06;
Matches 96; Conservative 67; Mismatches 152; Indels 95; Gaps 20;

```
QY 239 VWSHPVQDPVTGYVSNYGYQLVIAMGIPNSPT-----GDHNI 278
DB 44 IWDTPMLRD-FDGEIISVNGWCIIFTLTADRNTPQFDENGNYDITRDWEDRHRARI 102
QY 279 YLLNKGNDGDFSHRNAGSIFGTKEINVFQWSSAIV--VNDGTTIQLFFTSNDTSDYKL 337
DB 103 CYYWSRTCKD---WIFGRVMAEGVATPTREWAGTPTLLDRGDIDLIYTCVTPG---- 154
QY 338 NDORLATATNLNVDDNGVSIKSVNYQVLFEGDGFHYQTYEQFANGKDRNDYCLDRDP 397
DB 155 --ATIAKVRGKIVTSDQSVSLGFGQOQTSLSFSDAGTIIQTTEE-----NAFWNRDP 204
QY 398 H-VVQLENGDRYLVEAN-----GTEDYQSDDOIYNWANYGDDAFNKSFKLLNNKKD 452
DB 205 SPFIDRNDGKLYMLFEGNVAGPRGSHEITQ-----AEMG-----NVPPGYEDVGGAKY 252
QY 453 RELAGLANGALGILKLTNNQSKPKVEEYVSLVSTLMACDEVYKXKLGDKYKLFVSVT 508
DB 253 QA-----GCVG-LAVAKDLSSG-WQILPLITAVGNDQTERPHVVFQDGKYLFLFIS 304
```

QY 509 RVSRGSDRELTAQNTIVGDNVAMIGYVSDSLMGKKPLNNSGVVLTASVPANWRTATYS 568
Db 305 -----HKYTFADNLGTGDGV--YGFVSDKUTGTYTMMNSGLVL--GNPSSQPFQYYS 353
QY 569 YVAVPVAGHPDQVLITSYMSN-----KDFASGEGNYATWAPSFVLOINPD 613
Db 354 HYVMP-----NGLVTSFSDSPVPMKGKDYRIG---GTEAPTVKLLKGD 393

RESULT 11
B97033
uncharacterized protein, related to enterotoxins of other Clostridiales [Imported] - Cld
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97033
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
A: Bacteriol. 183, 4823-4836, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97033
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2817 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79053.1; PID:ql5023993; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1079

Query Match 4.9%; Score 203.5; DB 2; Length 2817;
Best Local Similarity 21.7%; Pred. No. 0.0078;
Matches 143; Conservative 92; Mismatches 211; Indels 213; Gaps 35;

QY 34 NNYNGVAEVT-----ERQA--NGOIGVGDKKIISANSNTTSGTNOESSA--TNNT 80
Db 66 NNLTKVQTNFASMPNTNPKATNNSKILVAPKLNQASSNEGTPPKQAIPVTNVT 125
QY 81 ENAVNESKNTNTENAVNE-----NKNTNT-----ENAVNENKNTNTEN 124
Db 126 DN-----KNTFNESSINNEAPIKDTSKTSTSAOTKGSNDNNIPSNNTSTWTKN 179
QY 125 D-----NSOLKLT-----NNOPSAATQANLKLNPQAKAVQNAKIDAGSLTDD 169
Db 180 ENPSNTDIKTEAPANAPKDTFNNQSDSALAKNALSNNLNLAADSSTSKVTSSN--- 235
QY 170 QINELKINFSKSAEGAKLTFKDLLEGIGNAIVKQDPQYAIYSNAKEIK-----NMPAT 224
Db 236 --NDAPKVNNTSTDKKASLNNDSDQD---WTKDGKKY--YVNGVQKGFQSIKSIY 288
QY 225 YTVD---AOTGKAHLVDVDSHPVQDPVGYVSNYMGYQLVIAMMGIPNSPTGDNHIIYL 281
Db 289 YFNDDSGMGTG-----W-----LKYSNSYIFDASGYML-----TGLQINGT 326
QY 282 YNKYGDNDFSHRNAGSIFGKTRTNVQFQWGSAGAIYNDVDTIQL--FFTSNDTSDYKLNDQ 340
Db 327 Y--YGFND-----DGKLLGLQAIN-----NNYYFNNDVGMQGWITCNDKSYFDDNG 374
QY 341 RLATATLNLNVDDNG-----VSIKSVNDYQVLFEG-----DGPHY----- 375
Db 375 VMQTGLVHNNKYGYFGNDGKLLTGLQINNNYTYFYFDSNGVMQTDWITIDGSKYFVSNG 434
QY 376 --OT-----YEOFANGKRDENDCYLDRPHVQLENGDRYLFVAN----- 414
Db 435 VMQTGLIISGYGYGFA-----NGKLLTG---LQVINGNSY-YFDYNGIRLVSRTWITI 484
QY 415 TGTEDYQSDQIY--NWANYGDDAFNTKSSFKLLNNKKDLRELALANGALGILKLTNNO 472
Db 485 DGKDYFNODGILTDNWINYDCKYFYISG-----VKQTGLQ 521
QY 473 SKPKVEEYVPLVSTLMACDEYXXLGDYKYLFSVTRVSRGSDRELTAQNTIVGDN--VA 531

Db 522 N-----IDGNYFFDSSGIMQTLQKIDGK--TYFQDNGIR 556
QY 532 MGYVSDSLMGKKYKPLNNSGVVLTASVPANWRTATYSYVAVPVAGHPDQVLITSYMSNK 590
Db 557 QIGWITYO--NKKYY--FNSDGSQMOTDLKIYSYSTSPYNYHYQYGFNDGKLLTGLQTIK 613

RESULT 12
AH2515
hypothetical protein alx7304 [Imported] - Anabaena sp. (strain PCC 7120) plasmid pCC
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AH2515
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irig
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2515
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4936 <KUR>
A:Cross-references: GB:BA000020; PIDN:BAW78388.1; PID:gl17135842; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr7304
A:Genome: plasmid

Query Match 4.9%; Score 203.5; DB 2; Length 4936;
Best Local Similarity 20.4%; Pred. No. 0.018;
Matches 185; Conservative 123; Mismatches 320; Indels 277; Gaps 45;

QY 6 KNMAVATLVASASILMGVVTAAHQVQVESNNY----NGVAFVNTERQANGOIGVDGKIISA 61
Db 413 QNTAILTIVNDLPPQGTINFNI-----NNYTVNENGTSINLVRGT---GSDGEV--\$ 461
QY 62 NSNTTSGSTNOESSATNTENAVNENKNTNTENAVNENKNTNTENAVNENKNTNN 121
Db 462 VLTLPDSGDTAGSDYNNLPITVTFANGETSKTN--LISONQGLFFDNDYVDNPNFSE 520
QY 122 TENDNSQLKLTNNEQPSAATQANLKLNPQAKAVQ---NAK-----I 161
Db 521 TK-DTTIELWAN---PTATRAS---TPETSGVNAFFNQYAIIPKQGLGLTGSTNDV 572
QY 162 DAG-----SLTDDQINELKINFSKSAEGAKLTFKDLLEGIGNAIVKQDPQYAIYS 213
Db 573 YAGISIGTNGVTISEHTLNYMPSVLVNTA-----LSGWNHIALVYENKTPKLYI 622
QY 214 NAKEIK--NMPATYTVDAQT-----GKMAHLVDVDSWPVDDV----- 249
Db 623 NQCFIKAGLTQIYIVHPSLFGGTSIROEDWFSKSIDDVRIWHKARTETEEKAGLNREL 682
QY 250 TGYVSNYMGYQLVIAMMGIPNSPTGDNHIIYLLYKYGDN---DFSHRNAGSIFGKTRN 306
Db 683 TGNESGLIG-----YWNFSINGNIVODLSTNNKNGTFFFGAQTSA 722
QY 307 VFOEWSGSAIVND-----GTQLFTSNDTSDYKLNQDLATATLNLNVDDNGYSIKSV 361
Db 723 GF---STSFINDINITYEPIETVNLTLT--NPTGGATLGTQK---TANLIVNDNDIAGTI 774
QY 362 D-----NYQVLFEGDGFHYQTEQFANGKRDENDCYLDRPHVQLENGDRYLVFANTGT 417
Db 775 QFSNANYAVNENGTAVNAVTILNR--TNGSD---GVVSVR-----INTNG-----TATAG 819
QY 418 EDYQSDDOIYWNANYGDDAFN-----IKSSFLLNNKKDLRELALANGAL-----GIL 466
Db 820 SDYNNSPITVNFAD--GENSKVTIPIIDDSILESNEISINLTLANPTNGATIGTQNSAVV 877
QY 467 KLTNNQSPKPY-----BEVYSPLVSTLMACDE-----VXX 496
Db 878 NIINDNLKPLTLVNTABQLTEGNTIOGTVTNRNTDTTEPLTVTLVNSDNTQITVPTVTI 937

148 -----LNQAAKAVONAKIDAG-----SLTDQINELNKNINFSKSAEKGAKLTFKDL 194
185 GSHKKNFAITVNGOVLYDENGALSSTSYSTFTQETTLNLTVDFTKNNAAYDSTRKASEFLV 244
195 EGIGNAIYKODPOYAIPYSNAKEIKNPNATYTVDAOTGKMAHLWDSPWPQDPVTVGVVS 254
245 DGYLTA-----DSWY-----RPEILEAGTWKATEKDFRLLMSW--WPKDQTOAVYL- 292
255 NYMGYQL-----VTAMGIPNSPTGDN-----HIYLLYKNKYGD----- 287
293 NYMTKALNGEBETKDVETIENSQASLAAQAQILQKIEVKIAANKSTDWLRQSTAEAFVKD 352
288 -----NDFS-----HWRNAGSIFCTKTNVFOEWSGS-----AIYN-DD 320
353 QDKNNINSESPGKEHFKGALLFVNSDST--KWANSDYRKLNQATATSYIKNHKIVNGSD 409
321 GTIQLTSTNDTSD-----YKLNQRLATATLNLNVDNDNGVSIKSVDNQVQV 366
410 GGYE-FLLSNDIDNSNPVQAEMLNQLYIFMNGQIVFGDKDKDAHFDGIRVDAVDNVSV 468
367 LFEGDGFHYQYEQFANGKORENDDYCL-----RDPHVY-----OLENGDR 407
469 -----DMQLVSSYKAAKYVNESARALANISILEAWSHNDPYVNEHNTAALSMNDGLR 524
408 YLVFE-----ANTGT-----EDYQSDDOIYNWANY----- 432
525 LSVHGLTRPTVNTKGTARNASMKDLINGYFGLSNRAEYTSVDQL-GFYATYLFVRAHDS 583
433 -----GGDDAF-----NIKSSFKLLNKKDRELAGLANGALGILKLTNNQ 472
584 EVQTVIADIISKIDPTDGTFTLQDKQAQFIYN-----ADMLKVDKEY 629
473 SKPKVEEYVPLVSTLMACDEVXXKLGDKY-----YLFSTVRSR----- 512
630 THSNIPAYALMLOTGCAATRY--VGDLYTDNGQYMAKSPIDQITLLKARPXYVAG 687
513 -----GSDRELTAQDNTVGDNVAMIGYVSDSLM-----GRYKPLNNGVW-LTA 556
688 GQTSYIHNLAGDGVSSAKDNKEY---LVSVRYGQDLSKTDTEGGKYG--RNSGMLTLIA 742
557 SVP-----ANWRATYSYIYAVPVAGHPDOVL-----ITSYMNKD-----FASGEG 597
743 NNPDKLADGETIVNW-----GAHRNQAYRPLLGTKEGIVSSLNDSDTKIYKYTDAQG 798
598 N-----YATWAFSFLVQINPDDTTVLA-----RAYNOGDWYWDSS 634
799 NLVFTADEIKGFKTVDMSGYLSVWVP-----VGATDQNVLAQPKSTKAYKEGDKVYSSSA 853
635 RNDNMLGLVKEGAAN 649
854 ALEAQ--VIYEGFSN 866
RESULT 15
B86748
hypothetical protein ykbc [imported] - Lactococcus lactis subsp. lactis (strain IL1403),
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: B86748
R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: B86748
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1093 <STO>
A:Cross-references: GB:AE005176; PID:g12723930; PIDN:AAK05084.1; GSPDB:GN00146
C:Genetics:
A:Gene: ykbc

Query Match 4.7%; Score 195.5; DB 2; Length 1093;
Best Local Similarity 19.5%; Pred No. 0.0054;
Matches 169; Conservative 122; Mismatches 236; Indels 341; Gaps 50;

QY 2 YKVKKNVAVTLVSASILMGV-VTAHADQVESNNYNGVAEVNTERQANGQIGVDGKIIS 60
Db 11 WSKGKIMLFSSGILLIFVIGSIGLTAQADEVPSN-----TETMTQ-----SSDVTTQSPVS 62
QY 61 ANSNTT-----SGSTNOESSATNTEN-AVNESKKNNTNENAVNENKNTNNTENAV----- 112
Db 63 PTVNSSLAASDSTTEEPVAVSNMNTAQVSSSTNLTSTESVTDTNPTQSNQNSIAISL 122
QY 113 -----VNE-----NKNTN-----NTENDNSOLKLTNNEOPSAATQANLKLNPQAAK 154
Db 123 TKASRESSINEPASTIANETTTADVTOTTAQAO--TSDFTTSLSESQCKPTNPQSK 181
QY 155 AVQ-----NAKIDAGSLTDDQINELNKNINFSKSAE-----KGA 187
Db 182 STEITNIQTVGAVDNNATGSAFVGDNITLQGDITDDNLLDPSLGHWSQOTQVIAIKGT 241
QY 188 KLTFKDLLEGIGNAIVKODPOYAIPYSNAKEIKNPNATYTV-DAQTGKMAH-----L 237
Db 242 ATGOLNQENFG---VQDGP--IIP-----ATTYMNWDGSGRITYVKGSLSGLDL 286
QY 238 DV-----WDSWPVQDPVTGYVS--NYMGYQLVIAAMGIPNSPTGDNHIYLLYKNKYGD 287
Db 287 DMIVTVASSDKDSQAANEGAEGIPQGLTFTGEONI-----AESGGSIVCLYN--GA 336
QY 288 NDFSHWRNAGSIFCTKTNVFOEWSGSAIYNDGTIQL-----FFTS--NDTSDYKLNQD 340
Db 337 NALS-----LIYQ-----IVKHTTEVPVVASFITDIDNAQGVQTNLA 376
QY 341 RLAT-----ATLNLNVDDNGVSIKSVNDYQVL-----FEGDGFHYQ--TYEQFANGKD 386
Db 377 NLVTILPOTTLNLAGDGTIYDAS-PNVPGLDGVASLPYGGYLGAGFVSEFYNYAYAPAE 435
QY 387 RENDDYC-----LRDPHVY-----OLENGDRYLVEANTGT 417
Db 436 RADDSYFFAAGVRYDLFGSALQAHMNTQIRQNFVNYDFEGHKIQETDHYLGF--TG- 491
QY 418 EDY-----QSD-----DOIYN----- 428
Db 492 QDYNLPIPTIKGFGVNLNTENDASKNPNVNLINLHNLPYYGNHNNHNNIYQGTATPSP 551
QY 429 ---WANYGGDDAFNIKSSFKLLNKKDRELA---GLANGALGIL-----KLTNNQSKPKV 477
Db 552 TIGYQIGNPEA---SITYTPVNNKGASSVTLPLIMAGGKTDIYENYNSVANNTKVPV 608
QY 478 -----EEVYSPLVSTLMACDEVXXKLGDKY--YLFV----- 507
Db 609 VSGVDYVYMAKNLIGITVDQAKELYSDL---YVASTLEKNPNPDGFADYLAGVHYGSDFG 665
QY 508 -----TRVSRGSD---RELTAKDNTI---VGDNVA-----MIGYVSDSLM 541
Db 666 SLIQKPLNALPNGLNYSSTNTQYKEIVEKSVTDLSHLGTSIANSYNQNGLTGILGDGLA 725
QY 542 GKYPKPLNNGSVTLTASVPANWRATY-----SYAVPVAGHPDQVLITSYMSNK---D 591
Db 726 G-----LPVRWGNPNVAGVFGGISFF-----BANNTKLMSNLSLIGD 761
QY 592 FASGEGNYATWAPSFVLQINPDDTTVL 619
Db 762 VLTNMGNETTKSGS--ADYNADMADAVIL 787

Search completed: September 26, 2002, 18:25:13
Job time: 92 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 18:24:20 ; Search time 25.32 Seconds
(without alignments)
764.023 Million cell updates/sec

Title: US-09-995-587A-11
Perfect score: 4149

Sequence: 1 MYKVKNAVATLVASILM.....LYAGSLALFGLAAIEKRA 792

Scoring table: BLOSUM62

Gap 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	649.5	15.7	455	2	US-08-870-827-3
2	267	6.4	578	4	US-09-503-172A-2
3	236.5	5.7	415	1	US-08-381-936-2
4	236.5	5.7	415	3	US-08-943-374-2
5	197	4.7	543	1	US-08-362-232-2
6	197	4.7	543	1	US-08-814-196-2
7	179.5	4.3	1183	2	US-08-447-031A-2
8	173	4.2	933	3	US-08-293-728-2
9	173	4.2	933	4	US-09-421-868-2
10	173	4.2	1004	4	US-09-268-347-30
11	173	4.2	2391	2	US-08-446-855A-2
12	173	4.2	2391	4	US-09-150-741-2
13	170.5	4.1	1612	1	US-08-169-927-2
14	170.5	4.1	941	4	US-09-336-447A-9
15	170	4.1	1098	1	US-08-409-995-2
16	170	4.1	1098	3	US-08-685-467-2
17	170	4.1	1098	4	US-09-377-155-32
18	170	4.1	1098	4	US-08-913-942-2
19	170	4.1	1098	4	US-09-669-974-32
20	170	4.1	1098	4	US-09-268-347-44
21	167.5	4.0	1104	4	US-09-268-347-28
22	167.5	4.0	1104	4	US-09-268-347-34
23	167.5	4.0	2411	4	US-09-268-347-36
24	166	4.0	1296	3	US-08-470-260-3
25	166	4.0	1296	3	US-08-471-491-3
26	166	4.0	1296	4	US-08-466-662-3
27	164.5	4.0	2123	4	US-08-968-685A-10

28	162.5	3.9	2354	4	US-09-268-347-47	Sequence 47, Appl
29	162	3.9	2353	4	US-09-377-155-33	Sequence 33, Appl
30	162	3.9	2353	4	US-08-913-942-4	Sequence 4, Appl
31	162	3.9	2353	4	US-09-669-974-33	Sequence 33, Appl
32	161	3.9	1577	2	US-08-793-824-2	Sequence 2, Appl
33	160	3.9	1475	3	US-09-007-999-2	Sequence 2, Appl
34	160	3.9	1475	3	US-09-210-361-2	Sequence 2, Appl
35	160	3.9	1912	1	US-08-409-995-4	Sequence 4, Appl
36	160	3.9	1912	3	US-08-685-467-4	Sequence 4, Appl
37	159.5	3.8	2314	4	US-09-268-347-49	Sequence 49, Appl
38	159	3.8	1430	3	US-09-008-172-2	Sequence 2, Appl
39	159	3.8	1430	4	US-09-210-361-6	Sequence 6, Appl
40	156.5	3.8	1529	2	US-08-728-470-10	Sequence 10, Appl
41	156.5	3.8	1529	2	US-08-719-641-10	Sequence 10, Appl
42	154.5	3.7	1600	2	US-08-617-697-10	Sequence 10, Appl
43	153	3.7	1702	4	US-08-296-791-5	Sequence 5, Appl
44	153	3.7	1702	5	PCT-US95-10661A-5	Sequence 5, Appl
45	151.5	3.7	1541	4	US-08-296-791-3	Sequence 3, Appl

ALIGNMENTS

```

RESULT 1
US-08-870-827-3
; Sequence 3, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 455 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-870-827-3

```

Query Match 15.7%; Score 649.5; DB 2; Length 455;
Best Local Similarity 37.8%; Pred. No. 7.3e-43;
Matches 158; Conservative 77; Mismatches 132; Indels 51; Gaps 17;
QY 203 KDDQVAIPTSNKAEIKNMPATVYDQAGKMAHLDVDSWPVQDPVTGYVSNMGVQLV 262
: ||: : || |||: : |||: : |||||: : |||:

Db 27 QNSPQKVPQFNASAIKNIDSACGYD-KSGLNLDLDVWDSWPLQN-ADGTAANYHGYHIV 84
QY 263 IAMGIPNSPTGDHNIYLLNKNYKQNDPFSHRNAGSIFGTKETNY-----FOEWSG 313
Db 85 SALAGDPKN-SDDTPLHLFYQKVGDTSIDSMKNAGRFEEDMKFVPNDPYLKYQTQEWGS 143
QY 314 SAIVNDGDTIQLFTSNDTSDYKLN-----DQRLATATNLNVDDNG-VSIKSYDN 363
Db 144 SAYLTKDQVRLFT-----DYSNPDGGTGGAGNOIISTAQVNLSPDAAATLKVGVGSD 198
QY 364 YQVLF-GDGFHYQTYEQFAN-GRDRENDYCLRDHPVVOLENGDRYLIVFEANTGTED-Y 420
Db 199 HKSVDGDDGTGYQVNIQOIEDEKWISSDHNHTRDPHYVE-DKCHKYLVFEANTGTDDY 257
QY 421 QSDQIYWNAYCGDDAFNFKSFKLNNKDKRELALANGALGILKLTNNQSPKPYEEV 480
Db 258 QGQSFNKAAYGGSVFFQNEKNKLLQSPK-KOIASLANGALGIVELADDYT---VKS 313
QY 481 YSPLVSTLMACDEV---XXKLGDKYVFSVTRVSRGSDRELTAKTNTIVGDNVAMIGYV 536
Db 314 MKPLVASNTVADEVERANIEFKMKNKWLFTRDSRGSKMTSGINDKD-----VYMLGPG 366
QY 537 SDSLMGKYPKLNNSGVVLTASVPANWRTATYSYAVPVAGHPD--QVLITSYMSKNDF 592
Db 367 GDSLNGPHNPINETGLVNLNMLPADLTHYSHCGIP---HPEGNNVVLTSYMTNRF 421

RESULT 2
US-09-503-172A-2
; Sequence 2, Application US/09503172A
; Patent No. 6284510
; GENERAL INFORMATION:
; APPLICANT: ITO, Tetsuya
; APPLICANT: FUJITA, Koki
; APPLICANT: HARA, Kozo
; APPLICANT: TONOSUKA, Takashi
; APPLICANT: SAKANO, Yoshiyuki
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE
; FILE REFERENCE: 10749-0001-0
; CURRENT APPLICATION NUMBER: US/09/503,172A
; CURRENT FILING DATE: 2000-02-14
; PRIOR FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Arthrobacter sp.
US-09-503-172A-2

Query Match 6.4%; Score 267; DB 4; Length 578;
Best Local Similarity 23.8%; Pred. No. 1.2e-12;
Matches 124; Conservative 82; Mismatches 196; Indels 118; Gaps 26;
QY 200 AIYKQDPQYAIYPSNAKETKMPATYTV-----DAQTGKMAHLVDVSWPQDPVTGYVS 254
Db 67 ADARQLQRMSDPTAPSRE-NSMPASVTMTPTVPQDFPDMSNEQVYVWDTPITDEDANQYS 125
QY 255 NYNGYQLVYIWMGIPNSPTGDNIHY-----LLYKNYK-----DNDSHWRNAGSIFGTK 303
Db 126 -VNGWEIIFSLVADRNLGDFDRHIVFAKIGYFYRPAAGVPAARPEN--GGWTYGLGVPEK 182
QY 304 ET-NVFO-----EWSGSAIVNDGCTIQLFT-----SNDTSYKLDORLATATLN 348
Db 183 VTQIFEDQSFSHTQWSSGARSYKNGEIKLFTDFVAFYNSGNTNKPYPDRIALSVGK 242
QY 349 LNVDDNGVSTKSYVDNYOVLFEGDFHYQTYEQFANGKDRENDDYCLDRPHVQ--LENGD 406
Db 243 VKANKKGVTLTGKNTVDLLQADGTYTQGAQ-----NEFFNFRDPTFEEDPAHPGE 294
QY 407 RYLIVFEANTGTEDYQSDQIYNNYNG---GD-----DAFNKSSFKLNNKDKREL 457

Db 295 TFMVFEGRSA---MORETATCNEADLGYROGDPYAEYVDDVNASGATYQIGN----- 343
QY 458 LANGALGILKLTNNQSPKVEEYVSPVSTLMACD-----EYXXKLGDKYVLSVYTR--- 509
Db 344 -----VGLAKAKNKOLTE--WEFLPILISANCVTQOTERPOIYFKDG-KSYLFTISHRGT 395
QY 510 VSRGSDRELTAKTNTIVGDNVAMIGYVSDSLMGKYKPLN-NSGVVLTASVPANW----- 562
Db 396 FAAGLD-----GPEGYVGVGDIRSDYQPLNGSGSLALGNPTNLNFLGGQPF 443
QY 563 -----RTATYSYIYAVPVAGHPDQVLITSYMSKNKDFASGEGNYATWAPFLVQI--N 611
Db 444 APDFNQHPGHFQAYSHYMP--GGLVQSFIDTIGTHDDFVRG---GTLAPTVMKMDIGVG 497
QY 612 PDDTTTTLARA--TNOGDWVDDSSRNNDNMLGVLKEGAANS 650
Db 498 GDPKTAVDYSYSGSEGLGWADIPANKHLFTNGKFGVAVS 537

RESULT 3
US-08-381-936-2
; Sequence 2, Application US/08381936
; Patent No. 5792923
; GENERAL INFORMATION:
; APPLICANT: ROBER, Manuela
; APPLICANT: GEIER, Gebhardt
; APPLICANT: GEIDER, Klaus
; APPLICANT: WILLMITZER, Lothar
; TITLE OF INVENTION: DNA sequences which lead to the
; TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
; TITLE OF INVENTION: these sequences as well as a process for preparing
; TITLE OF INVENTION: transgenic plants.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostiolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/381,936
; FILING DATE: 09-FEB-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93 02110
; FILING DATE: 09-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42270518
; FILING DATE: 08-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weillman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-381-936-2

Query Match 5.7%; Score 236.5; DB 1; Length 415;
Best Local Similarity 23.4%; Pred. No. 1.7e-10;
Matches 96; Conservative 67; Mismatches 152; Indels 95; Gaps 20;

QY 239 WDSWPVQDPVTGYVSNYMGVQLVIAMGIPNSPT-----GDNHI 278
DB 44 IWDTPMLRD-FDGEIISVNGWCIIIFLTADRNTDNPQFDENGNYDITRDWEDRGRARI 102
QY 279 YLLNKYGDNDFSHRNAGSIFGTKEINVOEWSGSAI-VNDGDTIQLFFTSNDTSYKL 337
DB 103 CYWYSRTGKD-----WIFGGRVMAEGVAPTTREWAGTPIILLNDRGIDILYTCVTPG--- 154
QY 338 NDQRLATATLNLVNDGNGSVKSDNYQVLFEGDGFHYQYEQFANGKRDNDYCLRDP 397
DB 155 --ATIARVGRKIVTSQSVLEGGFQQVTSLSADGTIYQTEEQ-----NAFWNRDP 204
QY 398 H-VVQLENGDRYLVEFANT-----GTEDYQSDQIYNWANYGGDDAFNIKSFKLNNKKD 452
DB 205 SPFIDRNDGKLYLMFEGNAGPRGSHEITQ-----AEMG-----NVPPGYEDVGGAKY 252
QY 453 RELAGLANGALGILKLTNNQSKPVEEYVSLVSTLMACDEVXXK-----LGDKYLLFSVT 508
DB 253 QA-----GCVG-LAVAKDLGSE-WQILPLITAVGVNDQTERPHFVFDGKYLLFTIS 304
QY 509 RVSRGSDRELTAKNITVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYS 568
DB 305 -----HKYTFADNLGTGPDGV--YGFVSKLGTPTPMNSSLGLVL--GNPSSQPFQYIS 353
QY 569 YYAVPVAGHPDQVLITSYMSN-----KDFASGEGNYATWAPSLVQINPD 613
DB 354 HYVMP-----NGLVTSFIDSVPWKGKDYRIG-----GTEAPTVKILLKGD 393

RESULT 4
US-08-943-374-2
; Sequence 2, Application US/08943374
; Patent No. 6028249
; GENERAL INFORMATION:
; APPLICANT: ROBER, Manuela
; APPLICANT: GEIER, Gebhardt
; APPLICANT: GEIDER, Klaus
; APPLICANT: WILLMITZER, Lothar
; TITLE OF INVENTION: DNA sequences which lead to the
; TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
; TITLE OF INVENTION: these sequences as well as a process for preparing
; TITLE OF INVENTION: transgenic plants.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,374
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/381,936
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42270618
; FILING DATE: 08-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mellman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-108

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-943-374-2

Query Match 5.7%; Score 236.5; DB 3; Length 415;
Best Local Similarity 23.4%; Pred. No. 1.7e-10;
Matches 96; Conservative 67; Mismatches 152; Indels 95; Gaps 20;

QY 239 WDSWPVQDPVTGYVSNYMGVQLVIAMGIPNSPT-----GDNHI 278
DB 44 IWDTPMLRD-FDGEIISVNGWCIIIFLTADRNTDNPQFDENGNYDITRDWEDRGRARI 102
QY 279 YLLNKYGDNDFSHRNAGSIFGTKEINVOEWSGSAI-VNDGDTIQLFFTSNDTSYKL 337
DB 103 CYWYSRTGKD-----WIFGGRVMAEGVAPTTREWAGTPIILLNDRGIDILYTCVTPG--- 154
QY 338 NDQRLATATLNLVNDGNGSVKSDNYQVLFEGDGFHYQYEQFANGKRDNDYCLRDP 397
DB 155 --ATIARVGRKIVTSQSVLEGGFQQVTSLSADGTIYQTEEQ-----NAFWNRDP 204
QY 398 H-VVQLENGDRYLVEFANT-----GTEDYQSDQIYNWANYGGDDAFNIKSFKLNNKKD 452
DB 205 SPFIDRNDGKLYLMFEGNAGPRGSHEITQ-----AEMG-----NVPPGYEDVGGAKY 252
QY 453 RELAGLANGALGILKLTNNQSKPVEEYVSLVSTLMACDEVXXK-----LGDKYLLFSVT 508
DB 253 QA-----GCVG-LAVAKDLGSE-WQILPLITAVGVNDQTERPHFVFDGKYLLFTIS 304
QY 509 RVSRGSDRELTAKNITVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYS 568
DB 305 -----HKYTFADNLGTGPDGV--YGFVSKLGTPTPMNSSLGLVL--GNPSSQPFQYIS 353
QY 569 YYAVPVAGHPDQVLITSYMSN-----KDFASGEGNYATWAPSLVQINPD 613
DB 354 HYVMP-----NGLVTSFIDSVPWKGKDYRIG-----GTEAPTVKILLKGD 393

RESULT 5
US-08-362-232-2
; Sequence 2, Application US/08362232
; Patent No. 5641667
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Fructosyltransferase Enzyme, Method
; TITLE OF INVENTION: For its Production and DNA Encoding the Enzyme.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,232
; FILING DATE: 22-December-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-362-232-2

Query Match 4.7%; Score 197; DB 1; Length 543;
Best Local Similarity 23.8%; Pred. No. 3.3e-07;
Matches 117; Conservative 78; Mismatches 175; Indels 122; Gaps 31;

QY 220 NMPATYVDAQTKMAHLD--VWDSWPVQDPVTGYVNMGYQLVIAMGIPNSPTG--D 275
Db 78 NIPADFPV-----INPDVWVWDTWTLIDKHADQFS--YNGWEVIFCLTADPNAGYGFDD 129
QY 276 NHY-----LLYNK-----YGDNDFSHRNAGSIFGKTETNVFQWMSGSAI 316
Db 130 RHVHARIGFFYRRACIPASRRPVNGWTYGGHLFPDGASAQVYAGQYTNQ--AEWSSGR 188
QY 317 VND--DGTIQLFFT-----SNDTSYKLN--DQRLATATL--NLNVDDNGVSIKSVQVQLF 368
Db 189 LMQIHGNTVSVFYTDVAFNRDANANNITPPQAIITQTLGRIHADFNHWFGTGTAHTPLL 248
QY 369 EGDGFHYQTYEQFANGKDRENDYCLDRPHVQV--LENGDRYLVEANT-----GTEDYQS 422
Db 249 QPDGVLYQ-----NG--AQNEFFNFRDPTFEDPKHPGVNMYVFEAGTAGQGVANCTE 300
QY 423 DDQIYNWANYGGDDAFNIKSSFKLLNNKKDRELA--GLANGALGILKLTNNQSKPKVEEV 480
Db 301 AD-----LGFRPNDP--NAETLQELVDSGAYYQKANIOLA-----IATDSLKWK---F 345
QY 481 YSPLVSTLMACDEVXXKL-----GDKYILFSV---TRVSRGSDRELTAKNITVGNVAMI 533
Db 346 LSPILISANCVNDQTERPQVYLHNGKYYIFTISHRTTFAAGVD-----GPD-GVY 393
QY 534 GYVSDSLMGYKPLNNSGVLTASVPANWRTA-----TYSYKAVPVAG 576
Db 394 GFVGDGIRSDQPM--NYGSLGTGNPTDLNTAAGTDFDPSDQNPRAFQSYSHYVMPGG- 451
QY 577 HPDQVLITSYMSNKKDFASGEGNYATWAPSLVQINPDPTTTVLARATNQGDWVDD--SS 634
Db 452 -----LVESFIDTVENRRG-----GTLAPTVRVRI--AQNASAVDLRYNGGLGGYGDIPAN 501
QY 635 RND-NMLGVLKE 645
Db 502 RADVNIAGFIQD 513

RESULT 6
; Sequence 2, Application US/08814196
; Patent No. 5731173
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Fructosyltransferase Enzyme, Method
; NUMBER OF INVENTIONS: For its Production and DNA Encoding the Enzyme.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York

```

```

; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,196
; FILING DATE: 10-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,232
; FILING DATE: 22-December-1994
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-814-196-2

```

```

Query Match 4.7%; Score 197; DB 1; Length 543;
Best Local Similarity 23.8%; Pred. No. 3.3e-07;
Matches 117; Conservative 78; Mismatches 175; Indels 122; Gaps 31;

```

```

QY 220 NMPATYVDAQTKMAHLD--VWDSWPVQDPVTGYVNMGYQLVIAMGIPNSPTG--D 275
Db 78 NIPADFPV-----INPDVWVWDTWTLIDKHADQFS--YNGWEVIFCLTADPNAGYGFDD 129
QY 276 NHY-----LLYNK-----YGDNDFSHRNAGSIFGKTETNVFQWMSGSAI 316
Db 130 RHVHARIGFFYRRACIPASRRPVNGWTYGGHLFPDGASAQVYAGQYTNQ--AEWSSGR 188
QY 317 VND--DGTIQLFFT-----SNDTSYKLN--DQRLATATL--NLNVDDNGVSIKSVQVQLF 368
Db 189 LMQIHGNTVSVFYTDVAFNRDANANNITPPQAIITQTLGRIHADFNHWFGTGTAHTPLL 248
QY 369 EGDGFHYQTYEQFANGKDRENDYCLDRPHVQV--LENGDRYLVEANT-----GTEDYQS 422
Db 249 QPDGVLYQ-----NG--AQNEFFNFRDPTFEDPKHPGVNMYVFEAGTAGQGVANCTE 300
QY 423 DDQIYNWANYGGDDAFNIKSSFKLLNNKKDRELA--GLANGALGILKLTNNQSKPKVEEV 480
Db 301 AD-----LGFRPNDP--NAETLQELVDSGAYYQKANIOLA-----IATDSLKWK---F 345
QY 481 YSPLVSTLMACDEVXXKL-----GDKYILFSV---TRVSRGSDRELTAKNITVGNVAMI 533
Db 346 LSPILISANCVNDQTERPQVYLHNGKYYIFTISHRTTFAAGVD-----GPD-GVY 393
QY 534 GYVSDSLMGYKPLNNSGVLTASVPANWRTA-----TYSYKAVPVAG 576
Db 394 GFVGDGIRSDQPM--NYGSLGTGNPTDLNTAAGTDFDPSDQNPRAFQSYSHYVMPGG- 451
QY 577 HPDQVLITSYMSNKKDFASGEGNYATWAPSLVQINPDPTTTVLARATNQGDWVDD--SS 634
Db 452 -----LVESFIDTVENRRG-----GTLAPTVRVRI--AQNASAVDLRYNGGLGGYGDIPAN 501
QY 635 RND-NMLGVLKE 645
Db 502 RADVNIAGFIQD 513

```

203	QY	KODPOVAIPYSNAKEIKNNPATYTVDAQTGKMAHLVDWDSHPVQDPVTGTVSYNSMGYQLV	266
383	Db	LPFGDYIL-----KEIE-APRPIITFDKK-----EYPTMKDITDNOGY----	419
263	QY	IAMMGIPNSPTGDNHIIYLLYKNYGNDFESHRNRNAGSIFGTKEITNFVQEWMSGSAIVNDGDT	322
420	Db	-----FTTIENAKAIEKTKDVSQAQVWEGTKV-----	447
323	QY	IQLFFTSNDSYKLNQRLATATMLNVDNGVCSIKVDNTQVLFEGDGFHYQYEQFA	382
448	Db	-----KPTIFVFLKYQ-----DDN-QNTTPVDKAEIKLELDGTTTKVTWSNLP	488
383	QY	NGKDRENDDYCLDRPHVVOLENG--RYLVFEANTGTEDYQSDDOQIYNWANYGGDDAFNI	440
489	Db	-----END-----KNGKAITYLVKEVNAQGEDTTTEGYTK-----	519
441	QY	KSSFLLNNKKKDELAGLANGALGILKLTNNOSKPKVEEIVSPL-----VSTLMACDEV	494
520	Db	ENGLVVTNTEKPIETTSISGEKVMDDK--DNQDGKRPEKVSYNLLANGEKVKTL-----DV	573
495	QY	XXKLGDKYLLFSVTRVSRSGDRELTAKDNTIVGDNVAMIGYVSD-----SLMGKYKPLNNS	550
574	Db	TSETNNKYEFKDLPKYDEGKKIEY-----TVTEHDYK--DYTTDINGTTITNKYTPGETS	626
551	QY	GVVLTASVPANWTATYYSYAVPVAGHPQOVLITSYMSNKDRFASGE-----GNYA-TW	602
627	Db	-----ATVTKNWDNNNQ-----DGRKPTIEIKVELYODGK--ATGKTAILNESNNHTHW	674
603	QY	-----APSELVQINPDTTTLARATN-----QGDWYVDDSS	634
675	Db	TGDEKAKGOQVKYTVBELTKVKGYTHVDNNDGNLIVTNKYTPETTSISGEKVDD--	732
635	QY	RNDNMLGLVKEGAANSAAALPGEMGKPYD-----WSLNRSPGLKLPHPQVPQPKIDQP-	687
733	Db	-KDNQDGKRPEKVSYNLLADGEKVKTLDTSETNW-----KYEFKDLPKYDEGK	780
688	QY	--DOQPSQNTKNVTPG-NGDKPAGKATPDNTNIDPSAQPSQNTNIDPSAQSGQNTKN	744
781	Db	KIETVTVTEHDVKQYTTDINGTTITNKYTPGET-----SATVTKN	819
745	QY	VTPNEKOGKN-TDAK-OLPQTGNKSG	769
820	Db	WDDNNQDGRPEIKVELYQDGKATG	846

RESULT 8
 US-08-293-728-2
 ; Sequence 2, Application US/08293728D
 ; Patent No. 6008341
 ; GENERAL INFORMATION:
 ; APPLICANT: Foster,
 ; APPLICANT: McDevitt, Damien L.
 ; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene
 ; FILE REFERENCE: 05344.105011
 ; CURRENT APPLICATION NUMBER: US/08/293.728D
 ; CURRENT FILING DATE: 1994-08-22
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 933
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 ; US-08-293-728-2

```

Query Match      4.2%  Score 173; DB 3; Length 933;
Best Local Similarity 20.5%  Prod.No.6.1.e-05;
Matches 162; Conservative 90; Mismatches 273; Indels 264; Gaps 40;

6 KNAVATVLVASILMGVYTAHADQVSNRYNGVAEVNTERQANGQIGVDGKIISANSNT 65
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 KSIQVAGVLTGTLGIFGLLSKEADASENSVQTQDSASNEKSN-----DSSVSGAAPT 68
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 66 --TSGSTNOESSATNTNENAV-----VNESKNTNNT--ENAVVNE--NKNNTNNTENA 111
 Db 69 DDTNVSDTKSSNTNNGSETVAQNPAQOQETQSSSTNATTEETPTVGEATTTTNOANTP 128
 QY 112 VVNEKNTNTENDNSOLKLTNNQPSAAATQANLKLNPQAAKAVQNAKIDAGSLTDQOI 171
 Db 129 ATTQSSNTNABELVNO-----TSNETTENDTNTVSSVSPQNSTNAENV-----STTQDTS 179
 QY 172 NELNKFNSAEGAKLTPEKLEGIGNAIVKODPQVAIPYSNA-----KE 217
 Db 180 TEATPSN--NESAPOSTDASNKD-----VVNQAVNTSAPRMAFSLAAVAADAPAGTD 231
 QY 218 IKNPATYTVDAOTGKMAHLVDWDSWPVQDPVTGYVSNYMGYQLVIAMMGIPNSPT--GDN 276
 Db 232 ITNQLTNTVTVGIDSGTTV-----YPHQ---AGYVKLNYGF-----SVPNSAVKGD 274
 QY 277 HIYLLYKNGDNDFSHRNAGSIFGTRETNTVQFWSGSAIVNDGDTIQLFTSNDTSDYK 336
 Db 275 F-----KITVPKELNL-----DPENVKKT----- 347
 QY 337 L-----NDQRLATATLNLVNDNGVSIKSDNYQVLFEGDGFHYQYEQFANGKDREND 391
 Db 348 -----GNVTLATGISTTANKTTLVDYKYG-----KYNLSI----- 380
 QY 512 RGSDELTAKTNTI-----VGDNVAMIGYVSDSLMGKYKPLNNSGVVLITASVPANMR 563
 Db 381 KGTIDQIDKNTNTYQRTIYVNPVSGDNV-----IAPVLTLGNLKPNTDSNALI-----DQ 429
 QY 564 TATYSYYAVPVAGHPDQVLITYSMSN--KDFASGEGNYATWAPS---FLVQIN--PDDTTTV 618
 Db 430 NTSIKVYKVDNAAD-----LSESYFVNPFENFEDVTNSVNTFFPNQYKVEFTPDQIIT 485
 QY 736 QXSGQNTKN 744
 Db 580 DSGSDSTSD 588

RESULT 9
 US-09-421-868-2
 ; Sequence 2, Application US/09421868
 ; Patent No. 6177084
 ; GENERAL INFORMATION:
 ; APPLICANT: Foster, Timothy J.
 ; APPLICANT: McDevitt, Damien L.
 ; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene
 ; FILE REFERENCE: 05344.105011
 ; CURRENT APPLICATION NUMBER: US/09/421,868
 ; PRIOR FILING DATE: 1999-10-19
 ; PRIOR FILING DATE: 08/293,728
 ; PRIOR FILING DATE: 1994-08-22
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: Patent ver. 2.0
 ; SEQ ID NO 2
 ; TYPE: PRT
 ; LENGTH: 933
 ; ORGANISM: Staphylococcus aureus

US-09-421-868-2
 Query Match 4.2%; Score 173; DB 4; Length 933;
 Best Local Similarity 20.5%; Pred. No. 6.le-05;
 Matches 162; Conservative 90; Mismatches 273; Indels 264; Gaps 40;
 QY 6 KNAVATLVASILMGVVTAAHADOVESNNYNGVAEYNTERQANGQIGVGGKIISANSNT 65
 Db 14 KSIGVASLVGTLLGFLGLSKKEADASENSVTSQSDASNESKSN-----DSSSVSAAPT 68
 QY 66 --TSGSTNOESSATNTNENAV-----VNESKNTNNT--ENAVVNE--NKNNTNNTENA 111
 Db 69 DDTNVSDTKSSNTNNGSETVAQNPAQOQETQSSSTNATTEETPTVGEATTTTNOANTP 128
 QY 112 VVNEKNTNTENDNSOLKLTNNQPSAAATQANLKLNPQAAKAVQNAKIDAGSLTDQOI 171
 Db 129 ATTQSSNTNABELVNO-----TSNETTENDTNTVSSVSPQNSTNAENV-----STTQDTS 179
 QY 172 NELNKFNSAEGAKLTPEKLEGIGNAIVKODPQVAIPYSNA-----KE 217
 Db 180 TEATPSN--NESAPOSTDASNKD-----VVNQAVNTSAPRMAFSLAAVAADAPAGTD 231
 QY 218 IKNPATYTVDAOTGKMAHLVDWDSWPVQDPVTGYVSNYMGYQLVIAMMGIPNSPT--GDN 276
 Db 232 ITNQLTNTVTVGIDSGTTV-----YPHQ---AGYVKLNYGF-----SVPNSAVKGD 274
 QY 277 HIYLLYKNGDNDFSHRNAGSIFGTRETNTVQFWSGSAIVNDGDTIQLFTSNDTSDYK 336
 Db 275 F-----KITVPKELNL-----DPENVKKT----- 347
 QY 337 L-----NDQRLATATLNLVNDNGVSIKSDNYQVLFEGDGFHYQYEQFANGKDREND 391
 Db 294 VPPIMAGDOVLA-----NGV-----IDSDGNVIYTFDYVNTKD--DVK 330
 QY 392 YCLRDPHVVOLENGDRYLVFEANTGTEDYQSDDIYWNANYGGDDAFNIKSFKLLNNKK 451
 Db 331 ATLTPAVI-----DPENVKKT----- 347
 QY 452 DRELAGLANGALGILKLTNNOSKPKVEEYSPVSTLMACDEVXXKLGDKYLFESVTRVS 511
 Db 348 -----GNVTLATGISTTANKTTLVDYKYG-----KYNLSI----- 380
 QY 512 RGSDELTAKTNTI-----VGDNVAMIGYVSDSLMGKYKPLNNSGVVLITASVPANMR 563
 Db 381 KGTIDQIDKNTNTYQRTIYVNPVSGDNV-----IAPVLTLGNLKPNTDSNALI-----DQ 429
 QY 564 TATYSYYAVPVAGHPDQVLITYSMSN--KDFASGEGNYATWAPS---FLVQIN--PDDTTTV 618
 Db 430 NTSIKVYKVDNAAD-----LSESYFVNPFENFEDVTNSVNTFFPNQYKVEFTPDQIIT 485
 QY 619 LARATNOGDWVWDDSSRNNDMLGVLKEGAANSALPGEWGKPVWDS---LINRSPGLGLK 675
 Db 486 PYIVVNGH--IDPNSKGDALRLSTLYG--YNSNII--W-RMSWDNEVAFNNGSGSDG 538
 QY 676 PHQPVPKIDQPDQOPSGONTKNVTPGNGDKPAGKATPDNTNIDPSAQPSGONTNIDPSA 735
 Db 539 IDKPVPV--EQPDE-----PGEIE-----PIPEDSDSDPGSD--SGSDSNSDSGS 579
 QY 736 QXSGQNTKN 744
 Db 580 DSGSDSTSD 588

RESULT 10
 US-09-268-347-30
 ; Sequence 30, Application US/09268347
 ; Patent No. 6335182
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
 ; FILE REFERENCE: 1038-860
 ; CURRENT APPLICATION NUMBER: US/09/268,347

;; CURRENT FILING DATE: 1999-03-16
;; NUMBER OF SEQ ID NOS: 54
;; SOFTWARE: PatentIn ver. 2.0
;; SEQ ID NO 30
;; LENGTH: 1004
;; TYPE: PRT
;; ORGANISM: Haemophilus Influenzae
US-09-268-347-30

Query Match 4.2%; Score 173; DB 4; Length 1004;
Best Local Similarity 20.6%; Pred. No. 6.9e-05;
Matches 191; Conservative 122; Mismatches 333; Indels 280; Gaps 46;

Qy 5 GKNWAVATL-----VSAILMGVVTAAHADOVE--SNNYNGV--AEVNTQRAQNG 51
Db 162 GKHTVFTLEKDLNKNVATVSKLSLG---ANGKNVDITSDTNGLKFAKPTNGQ-NGN 216
Qy 52 IGVDGKIISANTTSGSTNQESSATNTE-----NAVNVESK 89
Db 217 VHLNG-IASTLTDITGTT---KSATNGVDVQVHNRAASVADVLNAGWNIQNGASVDFV 272
Qy 90 NTNNTENAVVKNKNTNTENAVVKNKNTNTENDNSOLKLTNNEQPSAAATQANLKL 149
Db 273 NYTDVDFVNGLNTNVNTTDAHAK-KTIVRVDTGLPVQVTEGTEVVKVNGEYEA 331
Qy 150 POAAKAVONAKIDAGSLTDDQI-----NELKNKFNKSA---EKGAKLTFKLEGIGNAI 201
Db 332 KQGSADMDKKKGLAKTKVLSANGTNPVKISNVADGTEHTDAVSFKQKAL---- 387
Qy 202 VKQDPOYALPYSNAEIKMNPATYTVDAQTGMAHLWDWSDPVPQVPTGYVSNYNGYQL 261
Db 388 --QDKQVTLASNAY-----ANGSDADGKK-----GIQT-----LSNGLNFKF 424
Qy 262 VIAMGIPNSPTGDNHLYLLNKNYKNDPFSHRNAGSIFGTETNVFQWMSGSAIVNDG 321
Db 425 -----KSTGGE-----LLNIKAEND-----TFTFKKGSVQVGDG 456
Qy 322 --TIQ-----LFTSNDTSDYKLNDRLATATNLNVD-----NGVSIKS 360
Db 457 KATIQDGAKTTLGLVEASELVDLSNKLKGVGTGDTGTDGTTDLTVKSGDKVTLLKA 516
Qy 361 VDNQVLFECDFHYQTYEQFANGKRDNDYD-----LRDPHVVOLENGDRYL 410
Db 517 GNLKVKQEGTFTYALKDELTDVKSVEFKDTANGANGASTIKTKDGLTITPANGAG--A 574
Qy 411 FEANTG-----TED-----YQSDQIYNNWYGGDDAFN-IKSSFLLNNKKDREL 459
Db 575 AGNANTANTISVTKDGISAGNKAVNVVGLKFKGDFDPLTSSADNLTQYDNAYKGLT 634
Qy 460 N-----GALGILK-----LTNNSKPKVEEVSPLVSTLMACDEVXX 496
Db 635 NLDEKSKQKQTPVADNTAATVGLRGLGWVISADTKGELNKEYNAQVRN---ANEVXF 691
Qy 497 KLGDKYLLSVTVRVSGRSDELT---AKDNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVV 553
Db 692 KSGNGINVSQKT-LDNGT-REITFELAKD-----ENAIAPFGSGSKALR-----DNTVAI 738
Qy 554 LTASVPANRRTATYSYVAVPVAGHPQOVLITSYMSNK---DFASGEGNTATWAPSELVQI 610
Db 739 GTCNVVNAEKSAGF-----GDP-----NIEDKAGSYAFGDNDRIT----- 775
Qy 611 NPDDTTTVILARATNOGDWVDDSSRNDNMLG-----VLKEGAANSALPGEWKPDWMSL 665
Db 776 --SKNTFVLGNSVNA-----KRDANGNVLTEKEVVGKDGAKTKVTVPQALGETVENS 827
Qy 666 -----INRSPGLGLKPHOPQKIDOPQOPSGQNTKNVTPGNGDKPAGKATPNTNI 718
Db 828 YLGNASTATKDKGNLK-----SDGTAGNTTTTAGATGTVNGFAGATAHGAIVSV 875
Qy 719 DPSAQP-----SGQNTNIDPSAQ---XSGQNTKNVTPGNEKOGKNTD----- 757
Db 876 GASGEERRIQNVAAGEISATSDAINGSQLYAVAKVGTNLAGQVKNVKGKADAGTASALA 935

Qy 758 AKQLPOTGNKSGLAGLYAGSLALFG 783
Db 936 ASQLPQ-----ASMPGKSMVSIAG 954

RESULT 11
US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-855A-2

Query Match 4.2%; Score 173; DB 2; Length 2391;
Best Local Similarity 20.3%; Pred. No. 0.00028;
Matches 115; Conservative 93; Mismatches 227; Indels 132; Gaps 28;

Qy 31 VESNNGVAVETQRAQNGIGVDGKIISANSNTTSGST-----NOESSATNNTENAV 84
Db 1275 VKHLYNEVDV-DKDTQLH-----KENNNNNMNSGNVENCKLNKESYGYNNSSNCI 1326
Qy 85 VNESKNTN--NTENAVVNE---NKN-----TNNTENAVV--NENKNTNNT----- 122
Db 1327 -----NTNNINIENNICHDISINKNIKTINNNSNNSINNENVTNLNCVSRAGSHIY 1381
Qy 123 -----ENDNSOLKLTNNEQPSAATQANLKNPOAKAVQNAKIDAGSLTDDQINEL 174
Db 1382 GKBEKSTGSDTNTLSAQNSNNNFSCNNENNNKAN-----VDVNVLEND-TKKR 1429
Qy 175 NKNFSKSAKAKLTFKDLLEGIGNAIVKODPOYAIPIYSNAKEIKNMPATYTVDAOTGKM 234
Db 1430 EDINTTTFVEGQNSVNNKNKENSLLKGDDEEDIVVWNLKKNENSVNNVDCRKKDM 1489
Qy 235 AHLDVDWSWPQDPTVTVGSVN--YMGYQL---VIAMGIPNSPTGDNHLYLLNKNYKNDNF 290
Db 1490 -----DGKINDECKTYKKNKYKMDGLNNNIVDELNGTSHSTNDHLYLDNFTSDEEI 1543

Qy	291	SHWRNAGSIFGKTETNMFQEWGSAIVNDDGTIOLFFTSNDTSDYKLNQDORLATATILNL	350
Db	1544	GNKKNM-DWILSKEKSISNKNPCNSYYVDSVY-----NNEYKLNKKKELLIDNENLN	1594
Qy	351	VDDNGYSIKSDVNYQVLFEGDGFHYQTYEQFANGKDR-----ENDDYCLRD-----PHVV	400
Db	1595	-----DEYNNVNMCSNTNNAFAFVNGKDRNDNLEND--CIEKNMDHTYKHYN	1641
Qy	401	OLEN-----GDRYLVEANTGTEDYDSDDOIYNWANYGGDAFNKTSFKLLNKKKORELA	456
Db	1642	RLANRSTERNMMLMVNNEKSNHEKGBHR-RLGNLKNKE-----KNNEKNGKNKKOKNY	1596
Qy	457	GLANGALGILKLTNNQSPKVBSEVYSPILVSTLMACDEYXXKLGDPKY-----LFSVTRVS	511
Db	1697	HYVNH-----KRNEYNSNNIESKENNV-----DDINKK-----EYEDENDIYFFTHSS	1742
Qy	512	RGSDRELTAKDNTIVGDNVAMIGYVSD	538
Db	1743	QGNDDL-SNDNYLSSEELNTEYDOD	1768

```

RESULT 12
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbanoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

```

[illegible]

Db	1490	-----DCKNINDECKYTKKNYKDMGLNNINI	VELSNGSHSTNDHLYLDNFWNTSDEEI	1543
Qy	291	SHWRNAGSIFGCTKTNVFEQWSAISAVDDGTTIQ	LFFTSNDTSDYKLNDOORLATATLNLN	350
Db	1544	GNNKNM-DMYLSKEKSISKNKPNYSYVVD5VY	-----NNEYKNNKMKELIDNENLN	1594
Qy	351	VDDGVSISKVDNTQVILFEGDGPHYQYIEQFANG	KDR-----ENDDYCLRD-----PHVV	400
Db	1595	-----DEYNMNVNMCNSYNNASAFVNGKDRND	LENLND--CIENKMDHYTKHYN	1641
Qy	401	QLEN-----GDRLYFEANTGTEDYQSDQIYN	WYANGDDAFNIKSFKLNNKKDRELA	456
Db	1642	RLNNRRSTNERMMLWVNNKESNHEKGHR-RL	NGLNKKKE-----KNMEKNKGKNKKDKNY	1696
Qy	457	GLANGALGILKLTNNQSKPKVEEYVSPVSTL	MACDEVYXKLGDKGY-----LFSVTRVS	511
Db	1697	HYVNH-----KRNNEYSNNITESKFNKV	-----DOLNCL---EYTEDENDIYYFTHSS	1742
Qy	512	RGSDRELTAKDNTIVGDNVAMIGYVSD	538	
Db	1743	OGNNDGL-SNDNYSLSSELNTDEYDDO	1768	

RESULT 13
 US-08-169-927-2
 ; Sequence 2, Application US/08169927
 ; Patent No. 5783441
 ; GENERAL INFORMATION:
 ; APPLICANT: Carl, Mitchell
 ; APPLICANT: Dobson, Michael E.
 ; APPLICANT: Ching, Wei Mei
 ; APPLICANT: Dasch, Gregory A
 ; TITLE OF INVENTION: Gene and Protein Applicable to the
 ; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
 ; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Counsel, Naval Medical R & D Command
 ; STREET: Bldg 1, T-12, 8901 Wisconsin Ave.
 ; CITY: Bethesda
 ; STATE: MD
 ; COUNTRY: USA
 ; ZIP: 20889-5606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION NUMBER: US/08/169,927
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/742,128
 ; FILING DATE: 08/09/91
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Spevack, A. David
 ; REGISTRATION NUMBER: 24,743
 ; REFERENCE/DOCKET NUMBER: 75,976
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 295-6759
 ; TELEFAX: (301) 295-1022
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1612 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-169-927-2

Query Match 4.1%; Score 172; DB 1; Length 1612;
Best Local Similarity 19.5%; Pred. No. 0.00018;

Matches 200; Conservative 117; Mismatches 366; Indels 342; Gaps 47;

```
QY 4 VGKNAVATLVASILMGCVTAHADQVESNNYNGVAEVTNTER-----QANGQIG 53
Db 245 VSKNGA-AIEFNVTGLGG-----NLKGIIEINTAAVAGKLIISLGAANAVIG 291
QY 54 VD-----GKIIS--ANSNTTSG-----STNOESSAT-----NNTENAV 84
Db 292 TDNGAGRAAGFTVSDNGNAATISQVYAKNVIQSANAGGQVTEHIVDVGLGCTTTFK 351
QY 85 VNESKNTNTENAVNENKNTNNT-----ENAVNENKNTNTE-----123
Db 352 TADSK-----VIITENSFGSTNFGNLDQIVVPDTKILKGNFIDGVKNNGNTAGVIT 404
QY 124 -NDNSQLKTNNEOPSAAQTAN-----144
Db 405 FNANGALVASTDPNIAVTNINATEAGAGVVELSGIHIAELRLONGGSIFKLADGTWIN 464
QY 145 -----LKKLNPOAKAVQ--NAKIDAGSLTDDQIN-ELNKINFSKAERKAKLTFKDL 194
Db 465 GPVNONALMNNNAAGSIQDLSAIIITDIGNGVGNAALQHITLANDASKILADGANI 524
QY 195 EG--IGNAIVKODPOYAIPIYSNAKEIKWPATYIVDAQTGMALHDVWDSPVQD-PVTG 251
Db 525 IGANVGGAIHFQANGGTIKLNTQ--NNIVVNFOLDITDKTGVVDASSLTNNQTLTING 582
QY 252 YVSNYMGYQLVIAMGIPNSPT---GD-----NHILYLYNKYGDNDPSHWRNA 296
Db 583 SIGTVANTKTLAGNLGSSKTLNAGDVAINELVIENNGSVQLNHNHTILTKTINAAQ 642
QY 297 GSI-----FCTKETNY-----FOEMSGSAIV 317
Db 643 GQIIVAADPLNTTTLADGTNLGSAENPLSTIHEATRAANADSLNVGKGYLVANNITT 702
QY 318 NDDGTIOLFTSNDTS-----DYKLDORLATATNLNVDNDGCVTSKSDVNTQVL 367
Db 703 NDANVGSILHFRSGGTSIVSGTVGGQGHKLNNLIL-----DNGTTVK-----744
QY 368 FEGDGHYQYEQPAGKRENDYCLDRPHVQLENG--DRYLVFEANTGETDYQSDQ 425
Db 745 FLGD-----TTF-----NGGK-----IEGRSILQISNNTYTDHVESADNTCTLEFVNDP 790
QY 426 I-----YWNANG-----GDDAFNTKSSPKLLN-----NKKRELALANG-----461
Db 791 ITVTLNKGAYFGVLKQVLIISGPGNIVFENIGNVGIHVHGAANSISFENASLGTSLFLPS 850
QY 462 --ALGILKLTNNQSKPVEEYSPVSTLMACDEVXXK--LGDKYILFVSVTRYSRGSDR 516
Db 851 GTPLDVLTIKSTVNGTVDFNAPIV-VVSGIDSMINNGQIIGDK--NIIALSLSGDN 906
QY 517 ELTAKDNTIYGDNVAMITGYVSDSLMGKYKPLNNSG-VYLTASVPAN-----561
Db 907 SITVNANTLYS-----GIRTTKNQGTVTLSGGMENPFTIYGLGLENGSP 952
QY 562 -WRTATY-----SYAVPVAGHPDOVLITVSMKNDF-----ASGEGNYATWAP 604
Db 953 KLVQVTFETDNNLGSIIANNVTINDVTLTGGIAGTFDAKITLGSVNGNAV-----1007
QY 605 SFVLQINPDPTTVLARATNOG-----WVWDDSSRNDNMLVLKGAANSALPGE- 656
Db 1008 RFVDSFSDPRSMIVATQANKGVTVYLCNALVSNIGSLDTPVASVTRTGNDSGAGLQGN 1067
QY 657 WGPVPW-----SLINRSPGLGLKHPQVPQIDQPDQPSGQNTKNVTTPGNGDKPAGKA 711
Db 1068 YSQUIDFGTVNLILSNVILG-GGTTAINGEID-----LLTNLIIFANGSTWGDN 1118
QY 712 TPONTNIDPSAQSGQ---NTNIDPSAQSGQNTKNVTPGNERKGNWTDKAPQTCGNK- 767
Db 1119 TSISTTLNVSNGIGQVIAEDAQVNATTGTGTTIKIQDNANANFSQYATYLIQGGARF 1178
QY 768 SGLAG 772
Db 1179 NGTLG 1183
```

```
RESULT 14
US-09-336-447A-9
; Sequence 9, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-9

Query Match 4.1%; Score 170.5; DB 4; Length 941;
Best Local Similarity 20.7%; Pred. No. 9.8e-05;
Matches 190; Conservative 108; Mismatches 380; Indels 239; Gaps 41;

QY 4 VGKNAVATLVASILMGCVTAHADQVESNNYNGVAEVTNTERQANGQIGVDG-----KI 58
Db 111 ICKNSTV-----GGFTNEA-----MGEYSTVAG--GANNQAKGNTSTVGGGNGKA 155
QY 59 ISANSNTTSGSTNQ-----ESSATNTENAVVNESKNTNTNTENAVVNENKNTN 106
Db 156 IGNNSTVVGGSNNQAKGEHSTTAGCKNNOATGNGSFAAGVENKADAN--NAVALGNKNTI 213
QY 107 NTENAVVNENKNT-----NTENDNSQ---LKLTNNEQPSAATOANLKLN-- 149
Db 214 EGTNSVATGSSNNTVTKGKNVFILGNTNTENAQSGSVLLGNNTAGKAAATTVNNAEVNGL 273
QY 150 -----POAAKAYON--AKIDAGSLTDDQINELNKNFSK---SAEKGAKLTFDLEGIN 199
Db 274 TLENFAGASKANANNIGTVSGS-----ENNERQIVNVGAGQISATSTDVNGSQLHALAK 329
QY 200 AIVKODPOYAIPIYSNAKEIKNM-PATYTVDAQTGKMA-----HLDVWDSWPVQDPVTGY 252
Db 330 AVAK-----NKSDIKGLKGVKELDKVGLSRDINSLHDDVDAN---QDSIAKN 376
QY 253 VSNYMGYQLVTAMM---GIPNSPTGDNHIIYLYNKYGDNDPSHWRNAGSIFG-----301
Db 377 KADIKGLNKEVKELDKVGLSRDIGS-----LHDDVDADNQDSIAKNKADIKGLNKEVKE 431
QY 302 -TKETNVFOEWSGSAIVNDGDTIQLFFTSNDTSDYKLDORLATATLNLN---VDNGVS 357
Db 432 LDKEVGLSRDIGS--LHDDVATNOADIKAQNOADIKTLNENVEEELNLSGLRIDOKADI 489
QY 358 IKSVDNVQVLFEGDGHYQYEQFANG-----KDRENDYVCLDRPHVVQ 401
Db 490 DNNINNIYELAQQDOHSSDIKTLKNVVEEGLLDLSGLRIDOKADLTDKITLKN---N 545
QY 402 LENG-----DRYLVFEANTGETDYQSDQI---YWNANGYGGDAFNK 441
Db 546 VEEGLLDLSGLRIDOKADIKAQNOADIAQNOTDIQDLAAYNELQDYAQKQTEAIDLNA 605
QY 442 SS-----FKLLNNKKDRELALANGALGILKLTNNQSKPKVEYVSPVSTLMAC 491
Db 606 SSANTDRIATAELGAENKDKAQIA-----KAQANENKDI-----AK 643
QY 492 DEVXXKLGDY-----YLFVSVTRYSRGSDRELAKDNTIVGDNVAMIG-----534
Db 644 NQADIQLHDKKITNLGILHSMVAVRAGVNGTGVATNKADIKAQNOADIANNKNIYELAQQ 703
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 18:26:15 ; Search time 61.92 Seconds
(without alignments)
1420.711 Million cell updates/sec

Title: US-09-995-587A-11
Perfect score: 4149
Sequence: 1 MYKVGKNVAVTLVSASILM.....LYAGSLALFGLAAIEKRHA 792

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:
2:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:
3:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:
4:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:
5:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:
6:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:
7:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:
8:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:
9:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:
10:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:
11:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:
12:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:
13:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:
14:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:
15:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:
16:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:
17:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:
18:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:
19:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:
20:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:
21:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:
22:	/SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795.5	19.2	881	22	ABG21651 Novel human diago
2	795.5	19.2	893	22	ABG07167 Novel human diago
3	789.5	19.0	774	22	ABG05590 Novel human diago
4	789.5	19.0	774	22	ABG11754 Novel human diago
5	782	18.8	1027	22	ABG21645 Novel human diago
6	773.5	18.6	736	22	ABG04946 Novel human diago
7	762.5	18.4	789	22	ABG04946 Novel human diago
8	762.5	18.4	789	22	ABG28407 Novel human diago
9	760.5	18.3	527	22	ABG04307 Novel human diago
10	760.5	18.3	789	22	ABG23205 Novel human diago
11	760.5	18.3	823	22	ABG25769 Novel human diago

12	758.5	18.3	493	22	ABG21647 Novel human diago
13	758	18.3	546	22	ABG21832 Novel human diago
14	758	18.3	562	22	ABG21605 Novel human diago
15	758	18.3	1095	22	ABG21602 Novel human diago
16	758	18.3	1095	22	ABG21823 Novel human diago
17	755.5	18.2	855	22	ABG04500 Novel human diago
18	752.5	18.1	929	22	ABG23199 Novel human diago
19	751.5	18.1	472	12	AA10671 B.amyloliquefacien
20	744	17.9	797	22	ABG18136 Novel human diago
21	742.5	17.9	644	22	ABG06230 Novel human diago
22	742.5	17.9	913	22	ABG02461 Novel human diago
23	740.5	17.8	689	22	ABG00041 Novel human diago
24	740	17.8	786	22	ABG02257 Novel human diago
25	731	17.6	1551	22	ABG28409 Novel human diago
26	723	17.4	542	22	ABG21613 Novel human diago
27	690	16.6	657	22	ABG21612 Novel human diago
28	690	16.6	657	22	ABG21834 Novel human diago
29	668	16.1	914	22	ABG09624 Novel human diago
30	668	16.1	3048	22	ABG23791 Novel human diago
31	662.5	16.0	487	19	AAW48306 Bacillus sp. V230
32	654	15.8	778	22	ABG13422 Novel human diago
33	651.5	15.7	828	22	ABG24549 Novel human diago
34	651.5	15.7	928	22	ABG21821 Novel human diago
35	651.5	15.7	928	22	ABG22394 Novel human diago
36	651.5	15.7	928	22	ABG26921 Novel human diago
37	646.5	15.6	436	22	ABG04298 Novel human diago
38	646	15.6	495	22	ABG04320 Novel human diago
39	643.5	15.5	708	22	ABG25779 Novel human diago
40	643.5	15.5	708	22	ABG28401 Novel human diago
41	634.5	15.3	692	22	ABG21599 Novel human diago
42	633.5	15.3	893	22	ABG26977 Novel human diago
43	622.5	15.0	426	22	ABG04288 Novel human diago
44	595.5	14.4	981	22	ABG21856 Novel human diago
45	586	14.1	417	22	ABG04289 Novel human diago

ALIGNMENTS

RESULT 1
ABG21651
ID ABG21651 standard; Protein; 881 AA.
XX
AC ABG21651;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21642.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
XX N-PSDB; AAS85838.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

Db 307 lanvvtslanluseslnngdtsiqpedsaseitrafdtlakalnttdsspsladgid 366
 Qy 80 TENA-----VYVESKNTN--NTENAVVNEKNKNTNTENAVVNEKNK-----TNTNTENDNSQ 128
 Db 367 taggsihvlsrdqtpilievgepllsdthvfks-----iredngrsgktvhtegdmnm 422
 Qy 129 LKLTNNEQSAATQANLKKLPNOAKAVONAKIDAGSLTDDQINELNKINFSKSAEKAGAK 188
 Db 423 -----nikkivkqatvltfttallaggaqafakennqay-ketygvsh 466
 Qy 189 LTFKDELEGNAIVKODQYAIPIYNAKEIKNMPATYTVDAOTGKMAHLVDVWDSWPVQDP 248
 Db 467 itrhmlqpkq--dqnekypvqfdqstikn-----iesakg----ldvwdswplqn- 513
 Qy 249 VTGYVSNYNGYOLVIAMGIPNSPTGDHNYLLYNYKYGDNDFSHRNAGSIFGTRETNYF 308
 Db 514 adgtvaeyngyhvfalagspkd-addtsiymfykvgdswknagrvi--kdsdkf 570
 Qy 309 -----QWGSIAVNDGDIQIUFFTSNDTSYKLNDRATATLNLNVDNGVS 357
 Db 571 dandpilkdtqgewsafdsqkirlfyt--dysgkhygkqsttaqvnvksddtik 628
 Qy 358 IKSVDNYOVLFEGDGFHYQTYEQFAN-GKDRENDYCLRDHPHVOLENGDRYLVEANTG 416
 Db 629 lngvedhktifdgdktyqnvqgfidegnytsghntlrphye-dkghkylvfeantg 687
 Qy 417 TED-YOSDDQIYNWYNGDDAFNKSFKLNNKKRELALANGALGILKLTNNQSKP 475
 Db 688 tengygeeslnkayvggntfnfrkesqklqqsakkrd-aelangalgielnndyt-- 744
 Qy 476 KYEEVYSLVSTLMACDEV-----XXKLGDKYVLFVSVTRVSRGSDRELTAKTNTIYGDNVA 531
 Db 745 -lkkvmkplitsntvtdeieranvfmngkwyft---dsrgskmti-----dginndiy 796
 Qy 532 MGYVSDSLMGKYPKPLNNSGVVLTASVPANWRTATYSYAVPVAGHPDOVLITSYMSNKK 591
 Db 797 mlgvysnsltgykplnktglvqmglpndvftfshfapqa-kgnnvitsymtnrg 855
 Qy 592 FASGEGNTATWAPSLVQINPDQTTTTLARATNQG 626
 Db 856 ff--edkktafapsflmnkgnktsvkvnsilleqg 888

RESULT 3
 ABG05590
 ID ABG05590 standard; Protein; 774 AA.
 XX
 AC ABG05590;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #5581.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.

N-PSDB; AA569777.
 New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
 Claim 20; SEQ ID No 35949; 103pp; English.
 The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity, and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 Sequence 774 AA;
 Query Match 19.0%; Score 789.5; DB 22; Length 774;
 Best Local Similarity 34.3%; Pred. No. 8 4e-42;
 Matches 222; Conservative 100; Mismatches 220; Indels 105; Gaps 27;
 Qy 144 NLKKLPNOAKAVONAKIDAGSLTDDQINELNKINFSKSAEKAGLTFKDELEGNAIVK 203
 Db 4 nikkivkqatvltfttallaggaqafakennqay-ketygvshlrhmlqpkq--q 60
 Qy 204 QDPQYAIPIYNAKEIKNMPATYTVDAOTGKMAHLVDVWDSWPVQDPVTGYVSNYNGYOLVI 263
 Db 61 qnekypvqfdqstikn-----iesakg----ldvwdswplqn-adgtvaeyngyhvf 109
 Qy 264 AMGIPNSPTGDHNYLLYNYKYGDNDFSHRNAGSIFGTRETNYF-----QWGS 312
 Db 110 alagspkd-addtsiymfykvgdswknagrvi--kdsdkfandpilkdtqgews 166
 Qy 313 GSAIVNDGDIQIUFFTSNDTSYKLNDRATATLNLNVDNGVSIKSVNYQVLFEGDG 372
 Db 167 gsatftsdkirlfyt--dysgkhygkqsttaqvnvksddtiklngvedhktifdg 224
 Qy 373 FHYQTYEQFAN-GKDRENDYCLRDHPHVOLENGDRYLVEANTGTED-YOSDDQIYNWA 430
 Db 225 ktyqnvqgfidegnytsghntlrphye-dkghkylvfeantgengygeeslnka 283
 Qy 431 NYGDDAFNKSFKLNNKKRELALANGALGILKLTNNQSKPKVEEYSPVLTSLMA 490
 Db 284 yygggntfnfrkesqklqqsakkrd-aelangalgielnndyt---lkkvmkplitsntv 339
 Qy 491 CDEV-----XXKLGDKYVLFVSVTRVSRGSDRELTAKTNTIYGDNVAMIGYVSDSLMGKYP 546
 Db 340 tdeieranvfmngkwyft---dsrgskmti-----dginndiymlgyvsnsltgyk 392
 Qy 547 LNNSGVVLTASVPANWRTATYSYAVPVAGHPDOVLITSYMSNKKDFASGEGNTATWAPSF 606
 Db 393 lnktglvqmglpndvftfshfapqa-kgnnvitsymtnrgff--edkktafapsf 449
 Qy 607 LVQINPDQTTTTLARATNQGDVWDSRNDNMLG-----VLKEG-----AANSAALPG 655
 Db 450 lmnkgnktsvkvnsilleqgltwlqvakraqlgggsgtvlrvrvrielaasthiala- 508

QY 656 EWGKPVDMSLI--NRSPGLGLKHQPVPKIDPQQPSGQNTKNVT-----PGNGDKPA 708
:
D 509 --arhdwrcnrgyparg-----pagaeaanvtgdyvpvegrya 549
:
QY 709 G-----KATPDNTNIDPSAQPSGONTNIDPSA---QXSG 739
:
D 550 adrrrvygyrrgrpprnyageeeegsgsfddpatdrqfsg 596
:

RESULT 4
ABGI1754
ID ABGI1754 standard; Protein; 774 AA.
AC ABGI1754;
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #11745.
EX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
PN WO200175067-A2.
PP 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS75941.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 42113; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 774 AA;

[illegible]

Db 611 sqklqsakrdaelangelgilelnndyt---lkkvmpkplitsntvtdeleranvfk 666
QY 499 GKYYLFVSIVRSRSDRELTAQNTIVGDNVAMIGYVSDSLMGKYPKLNNSGVILTASV 558
Db 667 ngkwyfift---dsrgskmti---dginsndiymlygvnsitgpykplnktglvlqmg 719
QY 559 PANWRTATYSYAVPVAGHPDQVLITSYMSNKDFASGEGNYATWAPSLVQINPDOTT 618
Db 720 dndvftfytshavpqa-kgnnnvitsymtnrgff--edkktafapsflmnikgnktsv 776
QY 619 LARATNOG 626
Db 777 knsileqg 784

RESULT 8
ID ABG28407 standard; Protein; 789 AA.
XX AC ABG28407;
XX 18-FEB-2002 (first entry)
DT Novel human diagnostic protein #28398.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS92594.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 20; SEQ ID NO 58766; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 789 AA;
Query Match 18.4%; Score 762.5; DB 22; Length 789;
Best Local Similarity 36.3%; Pred. No. 4.5e-40;
Matches 199; Conservative 95; Mismatches 183; Indels 71; Gaps 21;
QY 101 ENKNTNTENAVVNEKN-----TNTTENDNSOLKLTNNQPSAQTANLKKLNPOAKA 155
Db 286 knvdlkgfepdsiredngrsqktvhtgdmn-----nikkivkqatvl 330
QY 156 VQNAKIDAGSLTDDQINELNKNFNSAEKAGLTKFDEGIGNAIVKDPQVAIPVNA 215
Db 331 tfttallagatqafakennqkay-ketygvshithrhmliqpkq--qqekyqvqfdd 387
QY 216 KEIKNMPATYTVDAQTGKMAHLWDVSWPQDPVTGYVSNMGYQLYVAMGIPNSPTGD 275
Db 388 stikn-----iesakg-----ldvwdswplqn-adgtvaeingyhvvfalagspkd-add 435
QY 276 NHYLLYKYGDNDFSHWRNAGSIFGKTETNVF-----QWSSGSAIVNDGDTIQ 324
Db 436 tslymfqkvqgdnstdwnagrff--kdsdkfdandpilkddtqewsgsatftsdgkir 493
QY 325 LEFTSNDTSDYKLNQDRLATATNLNVNDNGVSIKSDYVOLFEGDGHYQTYEQFAN- 383
Db 494 lfyt--dysgkygkqsltagvnsksddtLkingvedhktfddgdktyqnvqgfide 551
QY 384 GKRENDYCLDRPHVQLENGDRYLVFEANTGTE-YOSDDQIYNWANYGGDDAFNKS 442
Db 552 gnytsgdntlrldphyve-dkgkylvfeantgtengyggeeslfnkayvggtnfrke 610
QY 443 SPKLLNKKDRELALANGALGILTKLTNNOSKPKVEEVSPLVSTLMACDEV----XXKL 498
Db 611 sqklqsakrdaelangelgilelnndyt---lkkvmpkplitsntvtdeleranvfk 666
QY 499 GKYYLFVSIVRSRSDRELTAQNTIVGDNVAMIGYVSDSLMGKYPKLNNSGVILTASV 558
Db 667 ngkwyfift---dsrgskmti---dginsndiymlygvnsitgpykplnktglvlqmg 719
QY 559 PANWRTATYSYAVPVAGHPDQVLITSYMSNKDFASGEGNYATWAPSLVQINPDOTT 618
Db 720 dndvftfytshavpqa-kgnnnvitsymtnrgff--edkktafapsflmnikgnktsv 776
QY 619 LARATNOG 626
Db 777 knsileqg 784

RESULT 9
ABG04307
ID ABG04307 standard; Protein; 527 AA.
XX AC ABG04307;
XX 13-FEB-2002 (first entry)
DT Novel human diagnostic protein #4298.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.

```

PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS68494.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID No 34666; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 527 AA;
XX
Query Match 18.3%; Score 760.5; DB 22; Length 527;
Best Local Similarity 36.1%; Pred. No. 3.5e-40;
Matches 197; Conservative 95; Mismatches 178; Indels 75; Gaps 21;
QY 99 VNEKNNTNTNENAVNENKNTNTNTENDNSQLKLTNNQPSNATQANLKLNPQAKAVON 158
DB 36 iredrn-----grsqktvhtegdmnm-----nkkkivkqatvltft 71
QY 159 AKIDAGSLTDQINELNKNFNSAERGAULTFKDLEGIGNAIVKQDPOYAPYNAKEI 218
DB 72 tallagatqafakennqay-ketygvshltrhdmqlpkq--qqekyqvvpfdasti 128
QY 219 KNPATVTVDAQTGMAHLDDVMSWPQDPVTVGYVSNMGYQLVIAVMGIPNSPTGDNHI 278
DB 129 kn-----iesakg---ldwdswwlpqn-adgtvaeqngyhvvfalagspkd-addtsi 176
QY 279 YLLNKKYGDNDFSHWRNAGSIFGTKEINVF-----QEWSSGSAIVNDDGTIQLFF 327
DB 177 ymfqkygdnsidskwnagrvf--kdsdkdandpdkldqqtqswgsatfstdqkrlfy 234
QY 328 TSNDTSYKLNDRQLATATLNLNVDNDGVSITKSVDNTQVLFEFGFGFHYQVTEQFAN-GKD 386
DB 235 t--dyskghyqkqsltaqvnvksdtdlkingvredhktfdgqkyqnvqqfidegny 292
QY 387 RENDYCLRPVHVOLENGDRYLVEANTGTED-YQSDQIYNWANYGGDDAFNIKSFFK 445
DB 293 tsgdnhtlrphyye-dkghkylvfeantgtengyqgeeslfnkayygggnfrkesqk 351
QY 446 LLNNKKRELAGLALGILKLTNNOSKPKVEEYVPLVSTLMACDEV----XXKLGDK 501
DB 352 lqgskakrd-aelangelgielndyt---lkkvmkpltsntvtdeieranvfkmgk 407

```

```

QY 502 YLFSTVTRVSRGSDRELTAKDNTIVGDNVAMITGVYVSDLSMKYKPLNNSGVVLTASVPAN 561
DB 408 wylft---dsrgskmti-----dginsndiymlgvysnsltpgypkplnktglvlgmgldpn 460
QY 562 WRTATYSYIYAVPVAGHPDQVLITSYMSNKDFASGGNGYATWAPSLVQINPDPTTTVLAR 621
DB 461 dvtftyshfavpqa-kgnvvitsymtngiff-eddkatfapsflmnikgnktsvvnks 517
QY 622 ATNOG 626
DB 518 illegg 522

```

RESULT 10

ABG23205

ID ABG23205 standard; Protein; 789 AA.

XX AC ABG23205;

XX DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #23196.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR N-PSDB; AAS87392.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX Claim 20; SEQ ID No 53564; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

```

SQ Sequence 789 AA;
Query Match 18.3%; Score 760.5; DB 22; Length 789;
Best Local Similarity 36.1%; Pred. No. 6e-40;
Matches 197; Conservative 95; Mismatches 178; Indels 75; Gaps 21;

QY 99 VNEKNKNTNTEAVVNEKNKNTNTNTDNSQLKLTNNEQPSAATQANLKLNPQAQAVON 158
Db 298 i redn -----grsqktvhtegdmn-----nikkivkqatvlft 333
QY 159 AKIDAGSLTDQINELNKNINFSKSAEKGAKLTFKDLLEGNAIVKODPOYAIPYSNAKEI 218
Db 334 tallagatqafakennqkay-ketygvshitrhdmqlipkq--qqnekyyvqpfqdsti 390
QY 219 KNMPATYTVDAQTGKMAHLVDWDSWPVDPTVGYVSNYMGVQLVIAMGIPNSPTGDNIH 278
Db 391 kn-----iesakg----ldvwdswplqn-adgtvaeingyhvfalagspkd-addtsi 438
QY 279 YLLYKNYKGNDFSHWRNAGSIFGKTETNVF-----QEWSSAIVNDGDTIOLFF 327
Db 439 ymfqkvqgdnidskwnagrvt--kdsdkfdandpilkdtqewsgsatftsdgkirlfy 496
QY 328 TSNDTSYKLNDRQLATLNLNVDDNGSVISKSVDNYQVLFEGDGFHYQTYEQFAN-GRD 386
Db 497 t--dysgkhygkqslttagvnsksddtlkingvedhktifdgdktyqnvqgfidegny 554
QY 387 RENDDYCLDRPHVQLENGDRYLVFEANTGTED-YQSDDDQIYNWANYGGDDAFNITKSSPK 445
Db 555 tsgdhtlrdphyve-dkghkylvfeantgtengyqgeeslnkayygggttnfrkesqk 613
QY 446 LLNNKKDRELALAGLALGILKLTNNQSKPKVEEYVSLVSTLMACDEV---XXKLGDK 501
Db 614 lqgsakkrd-aelangalgilelnndyt---lkkvmpkpltsntvtdeieranvfkmmgk 669
QY 502 YLFSVTRVSRGSDRELTAKNNTIVGDNVAMIGYVSDSLMGYKPLNNSGVVLTAASVPAN 561
Db 670 wylft---dsrgskmti----dginsndiymlygvnsnltgpykplnktglvlqmgldpn 722
QY 562 WRTATYSYAVPVAGHPDQVLITVSNKDFASGEGNATWAPSELVQINPDPTTTLAR 621
Db 723 dvftlyshfavpqa-kgnnvitsymtnrgff--edkktafapsflmnknkgnktsvvkns 779
QY 622 ATNOG 626
Db 780 lileq 784

RESULT 11
ABG25769
ID ABG25769 standard; Protein; 823 AA.
XX AC
AC ABG25769;
XX DT
DT 18-FEB-2002 (first entry)
XX DE
DE Novel human diagnostic protein #25760.
XX KW
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS
OS Homo sapiens.
XX XX
XX WO200175067-A2.
XX PD
PD 11-OCT-2001.
XX PF
PF 30-MAR-2001; 2001WO-US08631.
XX PR
PR 31-MAR-2000; 2000US-0540217.
XX PR
PR 23-AUG-2000; 2000US-0649167.
XX XX
XX (HYSE-) HYSEQ INC.

```

XX
PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS89956.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 56128; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 823 AA;

Query Match 18.3%; Score 760.5; DB 22; Length 823;
Best Local Similarity 36.1%; Pred. No. 6.3e-40;
Matches 197; Conservative 95; Mismatches 178; Indels 75; Gaps 21;

QY 99 VNEKNKNTNTEAVVNEKNKNTNTNTDNSQLKLTNNEQPSAATQANLKLNPQAQAVON 158
Db 332 i redn -----grsqktvhtegdmn-----nikkivkqatvlft 367
QY 159 AKIDAGSLTDQINELNKNINFSKSAEKGAKLTFKDLLEGNAIVKODPOYAIPYSNAKEI 218
Db 368 tallagatqafakennqkay-ketygvshitrhdmqlipkq--qqnekyyvqpfqdsti 424
QY 219 KNMPATYTVDAQTGKMAHLVDWDSWPVDPTVGYVSNYMGVQLVIAMGIPNSPTGDNIH 278
Db 425 kn-----iesakg----ldvwdswplqn-adgtvaeingyhvfalagspkd-addtsi 472
QY 279 YLLYKNYKGNDFSHWRNAGSIFGKTETNVF-----QEWSSAIVNDGDTIOLFF 327
Db 473 ymfqkvqgdnidskwnagrvt--kdsdkfdandpilkdtqewsgsatftsdgkirlfy 530
QY 328 TSNDTSYKLNDRQLATLNLNVDDNGSVISKSVDNYQVLFEGDGFHYQTYEQFAN-GRD 386
Db 531 t--dysgkhygkqslttagvnsksddtlkingvedhktifdgdktyqnvqgfidegny 588
QY 387 RENDDYCLDRPHVQLENGDRYLVFEANTGTED-YQSDDDQIYNWANYGGDDAFNITKSSPK 445
Db 589 tsgdhtlrdphyve-dkghkylvfeantgtengyqgeeslnkayygggttnfrkesqk 647
QY 446 LLNNKKDRELALAGLALGILKLTNNQSKPKVEEYVSLVSTLMACDEV---XXKLGDK 501
Db 648 lqgsakkrd-aelangalgilelnndyt---lkkvmpkpltsntvtdeieranvfkmmgk 703
QY 502 YLFSVTRVSRGSDRELTAKNNTIVGDNVAMIGYVSDSLMGYKPLNNSGVVLTAASVPAN 561
Db 704 wylft---dsrgskmti----dginsndiymlygvnsnltgpykplnktglvlqmgldpn 756

QY	57	KIISANSNTTSGTNOESSATNTNENAVVNESKNTN-NTENAVVNNKNTNNTNENAVVNE	115
Db	26	klvgtkalsttg---kalrtlptakvfslpnlndfkvpasilprksiredn---gr	78
QY	116	NKNTNTNENDNSQLKLTNNEQPSAATQANLKLNPQAAKAVQNAKIDAGSLTDDQINELN	175
Db	79	sqtvthtegdmnm-----nikkivkqatvltfttallagatqafakenn	123
QY	176	KINFSAEKGAKLTFKDLEGNAIVKQDPOVAIPYSNAKEIKNNPATYTVDAQTGKMA	235
Db	124	qkay-ketygvshtrhdmliqpkq--qnekyqvqfddgstikn-----iesakg---	171
QY	236	HLVDWSPVQDPVTGYVSNMGYQVLIAMMGIPNSPTGDNHLYLLNYKYGNDNDFSHWRN	295
Db	172	ldvwdswplqn--adgtvaeingyhhvvalagapkd--addtsiymfyqvgdndsidwkn	228
QY	296	AGSIFGTGKETNVF-----QEWSGSAIVNDGDTIQLFPTSDTSDYKLNDRQRLAT	344
Db	229	agrvt--kdsdkfdandpilkddqtqewsgsatftsdkirifyt--dysgkhygksltt	284
QY	345	ATNLNVDNDNGVSIKSDYNDYQVLFEGDGFHYQTEQFAN-GKDRENDYCLDRPHVVQLE	403
Db	285	aqnvksdttklknrvvedhktifdgdgkyqnvqgfidegnytsgdntltlrphyye-d	343
QY	404	NGDRYLVEANTGTED-YQSDDOQIYNWANYGGDDAFNIKSFLLNKKDRELALANGA	462
Db	344	kghkylvfeantgtengyqgeesalfnkayvggntfrrkesqlqgsakkrd-aelanga	402
QY	463	LGILKLTNNQSKPKVEEYVSPVSTLMACDEV---XXKLGDKYLYLFSVTRVRSGRSDREL	518
Db	403	lgilelnndyt---lkkvmkpltsntvtdeieranvfkmgkwyft---dsrgskmti	456
QY	519	TAKDNTIYGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYSYAVPVAGHP	578
Db	457	----dginsndiymlygvsnsitgpykplnktglvlqmgldpndvttfshfayvpa-kg	511
QY	579	DOVLITSYMSNKNDFASGEGNYATWAPSLVQINPDPTTV	618
Db	512	nnvvtysmtnrgif--edkktafapsflmniknktstv	549
RESULT 15			
ABG21602	standard; Protein; 1095 AA.		
XX	AC	ABG21602;	
XX	AC	ABG21602;	
DT	18-FEB-2002	(first entry)	
DE	Novel human diagnostic protein #21593.		
XX	Human;	chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging;	diagnostic; genetic disorder.	
XX	Homo sapiens.		
OS			
XX	WO200175067-A2.		
PN			
XX	11-OCT-2001.		
PD			
XX	30-MAR-2001;	2001WO-US08631.	
PF			
XX	31-MAR-2000;	2000US-0540217.	
PR	23-AUG-2000;	2000US-0649167.	
XX	(HYSB-) HYSEQ INC.		
PA			
XX	Drmanac RT, Liu C, Tang YT;		
PI			
XX	WPI; 2001-639362/73.		
DR	N-PSDB; AAS85789.		

PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
XX	biodiversity		
PS	Claim 20; SEQ ID No 51961; 103pp; English.		
XX	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. ABG00010-ABG3037 represent novel human		
CC	diagnostic amino acid sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
SQ	Sequence	1095 AA;	
Query Match 18.3%; Score 758; DB 22; Length 1095;			
Best Local Similarity 34.7%; Pred. No. 1.3e-39;			
Matches 201; Conservative 106; Mismatches 199; Indels 74; Gaps 23;			
QY	57	KIISANSNTTSGTNOESSATNTNENAVVNESKNTN-NTENAVVNNKNTNNTNENAVVNE	115
Db	559	klvgtkalsttg---kalrtlptakvfslpnlndfkvpasilprksiredn---gr	611
QY	116	NKNTNTNENDNSQLKLTNNEQPSAATQANLKLNPQAAKAVQNAKIDAGSLTDDQINELN	175
Db	612	sqtvthtegdmnm-----nikkivkqatvltfttallagatqafakenn	656
QY	176	KINFSAEKGAKLTFKDLEGNAIVKQDPOVAIPYSNAKEIKNNPATYTVDAQTGKMA	235
Db	657	qkay-ketygvshtrhdmliqpkq--qnekyqvqfddgstikn-----iesakg---	704
QY	236	HLVDWSPVQDPVTGYVSNMGYQVLIAMMGIPNSPTGDNHLYLLNYKYGNDNDFSHWRN	295
Db	705	ldvwdswplqn--adgtvaeingyhhvvalagapkd--addtsiymfyqvgdndsidwkn	761
QY	296	AGSIFGTGKETNVF-----QEWSGSAIVNDGDTIQLFPTSDTSDYKLNDRQRLAT	344
Db	762	agrvt--kdsdkfdandpilkddqtqewsgsatftsdkirifyt--dysgkhygksltt	817
QY	345	ATNLNVDNDNGVSIKSDYNDYQVLFEGDGFHYQTEQFAN-GKDRENDYCLDRPHVVQLE	403
Db	818	aqnvksdttklknrvvedhktifdgdgkyqnvqgfidegnytsgdntltlrphyye-d	876
QY	404	NGDRYLVEANTGTED-YQSDDOQIYNWANYGGDDAFNIKSFLLNKKDRELALANGA	462
Db	877	kghkylvfeantgtengyqgeesalfnkayvggntfrrkesqlqgsakkrd-aelanga	935
QY	463	LGILKLTNNQSKPKVEEYVSPVSTLMACDEV---XXKLGDKYLYLFSVTRVRSGRSDREL	518
Db	936	lgilelnndyt---lkkvmkpltsntvtdeieranvfkmgkwyft---dsrgskmti	989
QY	519	TAKDNTIYGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYSYAVPVAGHP	578
Db	990	----dginsndiymlygvsnsitgpykplnktglvlqmgldpndvttfshfayvpa-kg	1044
QY	579	DOVLITSYMSNKNDFASGEGNYATWAPSLVQINPDPTTV	618

Db 1045 nnvitsyntnrgff--edkktafapsflmnikgnktsvv 1082
: |:|||||: |: | |:|||||: |: |: |

Search completed: September 26, 2002, 18:26:20
Job time: 159 sec

Db 27 QNSPOKVPQFNASAIKNIDSAGYD-KSGNLIDLVDWSWPLQN-ADGTAANYGHVHIV 84
Qy 263 IAMGIPNSPTGDNHLYLLYKNYKGDNDPSHWRNAGSIFGTKEYNV-----FQWWSG 313
Db 85 SALAGDPKN-SDTPUHLFYQKVGDTSIDSWKNAGRVFEDMDKFPVNDPLYKYQTQWMSG 143
Qy 314 SAIVNDGDTIQLFFTSNDTSYKLN-----DQRLATATLNLNVDDNG-VSIKSDVN 363
Db 144 SATLTKDGOVRLFTY-----DYSNPEDDGTGAGNOIISTAQVNLQSDPAATLKVDCVSD 198
Qy 364 QYVLEP-GDGFHYQTYEQFAN-GKDRENDYCLDRPHVVOLENGDRYLVFEANTGTED-Y 420
Db 199 HKSVDGDDGTQYQIQFIDEGKWSGDNHTLRDPHYVE-DKGHKLVFEANTGTIDGY 257
Qy 421 QSDDOIYNWANYGGDDAFNIKSSFKLLNNKDBRELALGALGILKLTNNQSKPKVEEV 480
Db 258 QGDQSFNNKAYYGGSDVFFQNEKNKLQSPK-KQIASLANGALGIVELADDDY---VKSV 313
Qy 481 YSPLVSTLMACDEV---XXKLCCKYILFSTVRVSRGSDRELTAKTIVGDVNAVIGYV 536
Db 314 MKPLVASNTVADEVERANIFKNNKWLFTDSRGSKWTSDDGINDKD-----VYMLGPG 366
Qy 537 SDSLGMKYPLNNSGVVLTASVPANWRTATSYAVPVAGHPD--QVLITSYMSNKDF 592
Db 367 GDSLNGPHNPINETGLVLNNLDPADLTHTYSHCGIP---HPEGNNVLTSYMTNRGF 421

RESULT 2
US-09-503-172A-2
; Sequence 2, Application US/09503172A
; Patent No. 6284510
; GENERAL INFORMATION:
; APPLICANT: ITO, Tetsuya
; APPLICANT: FUJITA, Koki
; APPLICANT: HARA, Kozo
; APPLICANT: TOMOZUKA, Takashi
; APPLICANT: SAKANO, Yoshiyuki
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE
; FILE REFERENCE: 10749-0001-0
; CURRENT APPLICATION NUMBER: US/09/503,172A
; CURRENT FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 160416/1999
; PRIOR FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Arthrobacter sp.
US-09-503-172A-2

Query Match 6.4%; Score 267; DB 4; Length 578;
Best Local Similarity 23.8%; Pred. No. 1.2e-12;
Matches 124; Conservative 82; Mismatches 196; Indels 118; Gaps 26;
Qy 200 ATVKODPOAIPYSNAKETKNPATYTV-----DAQTGKMAHLVDWDSWPDVPTGYVS 254
Db 67 ADARQLQMSDTPASRE-NSMPASVTMTPTVPQDFPDSMSNEQVWVWDTWPLTDEANQYS 125
Qy 255 NYMGQYLVIAMMGPNSPTGDNHIIY-----LLYKNYK-----DNDFSHWRNAGSIFGPK 303
Db 126 -VNGWEIIFSLVADNLGDFDRHVFAGIKGYFYPAGVPAERPEN--GGWTYGLLVFREG 182
Qy 304 ET-VWFQ-----EWGSAIVNDGGTIQLFFT-----SNDTSYKLNQRLATATLN 348
Db 183 VTGQIFEDQSFQSHQTSQSGSARVSKNGEIKLFFTDVAFYRNSDGTNRIKPYDPRIALSVCK 242
Qy 349 LNVDDNGVSIKSDVNYQVLFEGDGHYQTYEQFANGKORENDYCLDRPHVVQ--LENGD 406
Db 243 VKANKKGVTLTGFKNVKTLQADGTYQYTGAGQ-----NEFFNFRDPTFTEDPAHGE 294
Qy 407 RYLVEFANTGTEDYQSDQIYNWANYG---GD-----DAFNKSSFKLLNNKKDREL 457

Db 295 TMVFEGNSA---MORETATCNEADLGYRGDPYAETVDDVNASGATYQIGN----- 343
Qy 458 LANGALGILKLTNNQSKPKVEEVYSPVSTLMACD-----EVXXKLGDKYYLFSVTR--- 509
Db 344 -----VGLAKAKNKQLTE--WEFLPPLILSANCYVTDQTERPQIYFKDG-KSLFTLISHRGT 395
Qy 510 VSRGSDRELTAKTNTIVGDNVAMIGYVSDSLMGKYKPLN-NSGVVLTASVPANW----- 562
Db 396 FAAGLD-----GPEGVYGFVGDGIRSDYQPLNGSGSLALGNPTNLNLFGLGGQPF 443
Qy 563 -----RTATSYAVPVAGHPDQVLTITSYMSNKDFASGEGNYATWAPSLVQI--N 611
Db 444 APFNOHPGHFQAYSHYVMP--GGLVQSFDITGTHDDFVRG-----GTLAPTVMKDIGV 497
Qy 612 PDDTTTTLARA-TNQGDWVDDSSRNDNMLCVLKEGAANS 650
Db 498 GDPTKTAVDYSGSEGLGGWADIPANKHLFTNGKFGVAVS 537

RESULT 3
US-08-381-936-2
; Sequence 2, Application US/08381936
; Patent No. 5792923
; GENERAL INFORMATION:
; APPLICANT: ROBER, Manuela
; APPLICANT: GEIER, Gebhardt
; APPLICANT: GEIDER, Klaus
; APPLICANT: WILLMITZER, Lothar
; TITLE OF INVENTION: DNA sequences which lead to the
; TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
; TITLE OF INVENTION: these sequences as well as a process for preparing
; TITLE OF INVENTION: transgenic plants.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/381,936
; FILING DATE: 09-FEB-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93 02110
; FILING DATE: 09-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42270618
; FILING DATE: 08-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-381-936-2


```

; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.0 for DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,196
; FILING DATE: 10-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/362,232
; FILING DATE: 22-December-1994
; APPLICATION NUMBER: CU 125/93
; FILING DATE: 23-December-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; US-08-814-196-2

Query Match 4.7%; Score 197; DB 1; Length 543;
Best Local Similarity 23.8%; Pred. No. 3.3e-07;
Matches 117; Conservative 78; Mismatches 175; Indels 122; Gaps

Qy 220 NMPATYTVDAQTGRMAHLD---VNDSDHPVDPTGYGVNTNGYQLVTAMGIPNSPTG--D 275
   :|||: : |||: | : |||: | : |||: | : |||: | : |||: | : |||: | : |||: |
Db 78 NIPADFPV-----INPDVVMWDTLIDKHADQFS-YGNVEVIFCLTADPNAGYGFD 129

Qy 276 NHYY----LLYNK-----YGDNDFSHRNAGSIFGTKETNVFOEWSGSAI 316
   :||: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 RHVFARTGFYRRRAGIPASRRPVNGWTYGGHLPPDGASAQYVAGQTYNQ-AEWSGSSR 188

Qy 317 VND--DGTIOLFFT-----SNDTSYKLN-DQRLATATL-NLNYDDNGVSIXSVDNVQVLF 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 LMQIHGNTSVSYFTDVAFNRDANANNITPQAIIOTGLRIHADFNHVWFTGTHTPL 248

Qy 369 EGDGFHQVTEQFANGKDRENDYCYLRDPHVQ--LENGDRYLVEFANT-----GTEDYQS 422
   : || || : : : : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 249 QPDGVLXQ-----NG--AQNEFFNFDPFTFEKPFGVNYVWFEGNTAGQGRCVANCTE 300

Qy 423 DDQLYNNANYGGDDAFNIKSFSEKLLNNKKDRELA--GLANGALGIKLNTNQSKPKVEEV 480
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 AD-----LGFRPNP-NAETLQEVLDSGAYYQKANICLA-----IATDSTLSKWK---F 345

Qy 481 YSPLVSTLMACDEVXXKL-----GDKYILFSV--TRYSRGSDRELTAKDNTIVGDNVAMI 533
   ||| : : : : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 346 LSPILISANCVNQDTERPQVYLHNKGYYIFTISHRTFAAGVD-----GPD-GVY 393

Qy 534 GYVSDSLMGKYKPLNNSGVLTASVPANWRTA-----TYSYAVYPVAG 576
   :||: : : : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 394 GFVGDCIRSDQPIM-NVSGSLTWGNPNDLTNAGCTDFDPSDQNPRAFQSYSHVMPGG-- 451

Qy 577 HPDQVLIITSYMNKDFASGEGNYATWAPSELVQINPDTTTVTLARATNQCGDWWD--SS 634
   : : : : : : : : : : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 452 -----LVESFIDTVENRRG-----GTLAPTAVRRI-AQNASADVRLRYNGGLGGYGDIPAN 501

Qy 635 RND-NMLGLVKE 645
   | | | : : : :
Db 502 RADVNIAGFIQD 513

```

```

7
US-08-447-031A-2
; Sequence 2, Application US/08447031A
; Patent No. 5851794
; GENERAL INFORMATION:
; APPLICANT: GUSS, Bengt
; APPLICANT: HOOK, Magnus
; APPLICANT: JONSSON, Hans
; APPLICANT: LINDBERG, Martin
; APPLICANT: PATTI, Joseph
; APPLICANT: SIGNAS, Christer
; APPLICANT: SWITALSKI, Lech
; TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
; TITLE OF INVENTION: ITS PREPARATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/447,031A
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/861,804
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/SE91/00707
; FILING DATE: 22-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9003374-7
; FILING DATE: 22-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 012889-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-031A-2

```


[illegible]

QY 291 SHWRNAGSIFGTKETNVFQWSSSAIVNDGTTQLFFTSNDTSDYKLNDRORLATATNLN 350
Db 1544 GNNKNN-DWLSKEKSISNKNPCNSYVVDSVY-----NNEYKINKMKELIDNENLN 1594
QY 351 VDDNGVSIKSDNYQVLFEGDGFHYQTYEQFANGKDR-----ENDDYCLRD-----PHVV 400
Db 1595 -----DEYNNVNMNCSNYNNASAFVNGKDRNDNLEND--CIEKNMDHTYKHYN 1641
QY 401 QLEN-----GDRYLVEANTGTEDYQSDDOIYNWANYGGDDAFNIKSSFKLLNNKKORELA 456
Db 1642 RNNRSTNRMLAYNNKESNHEKGHR--RNLNKKNKE-----KMEKNKGNKKKKY 1696
QY 457 GIANGALGLKLTNNQSKPKVEEVSPLVSTLMACDEVXXKLGDKY-----LFSVTRVS 511
Db 1697 HYVNH-----KRNEYSNNIESKFNYY-----DDINKK---EYEDENDIYFTTHSS 1742
QY 512 RGSDELTAKTNTIVGDNVAMIGYVSD 538
Db 1743 QGNNDL-SNDNYLSSEELNTDEYDD 1768

RESULT 12

US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

Query Match 4.2%; Score 173; DB 4; Length 2391;
Best Local Similarity 20.3%; Pred. No. 0.00028;
Matches 115; Conservative 93; Mismatches 227; Indels 132; Gaps 28;
QY 31 VESNNYNGVAEVTNERQANGQIGVDGKIISANTSGT-----NOESSATNNATENAV 84
Db 1275 VKNHLYNEVD-DKDTQLH-----KENNNNNNSCNVENCKLNKESGYNNSSNCI 1326
QY 85 VNESKNTN--NTENAVNE-----KKN-----TNTENAVV--NENKNTNNT----- 122
Db 1327 -----NTNNIENNICHDISINKIKVTINNNSNISNNENVTNLCVSRAGSHHIY 1381
QY 123 -----ENDNSOLKLTNNQPSAATOANLKNLPQAKAVQNAKIDAGSITDDQINEL 174
Db 1382 GKEKSIGSDDTNLSAQSNNNSNENNNKNN-----VDVNVLEND--TKKR 1429
QY 175 NKINFSKAEKALTKFDLEGNAIVKQDPQYAIPTYSNAKEIKNNMPATYTVDAQTGKM 234
Db 1430 EDINTTTFMEGQNSVINNNKKNSSLLKGDDEEDIVVNLKKNKNNYSVINNVDCRKKDM 1489
QY 235 AHLQVDSWPQDPVTGVSVN-YMGYQL---VIAMMGIPNPTGDNHLYLYNKGDNDF 290

Db 1490 -----DGKNINDECKTYKKNKYKMDGLNNNIYDELNGTSHSTNDHLYLONFNTSDEEI 1543
QY 291 SHWRNAGSIFGTKETNVFQWSSSAIVNDGTTQLFFTSNDTSDYKLNDRORLATATNLN 350
Db 1544 GNNKNN-DWLSKEKSISNKNPCNSYVVDSVY-----NNEYKINKMKELIDNENLN 1594
QY 351 VDDNGVSIKSDNYQVLFEGDGFHYQTYEQFANGKDR-----ENDDYCLRD-----PHVV 400
Db 1595 -----DEYNNVNMNCSNYNNASAFVNGKDRNDNLEND--CIEKNMDHTYKHYN 1641
QY 401 QLEN-----GDRYLVEANTGTEDYQSDDOIYNWANYGGDDAFNIKSSFKLLNNKKORELA 456
Db 1642 RNNRSTNRMLAYNNKESNHEKGHR--RNLNKKNKE-----KMEKNKGNKKKKY 1696
QY 457 GIANGALGLKLTNNQSKPKVEEVSPLVSTLMACDEVXXKLGDKY-----LFSVTRVS 511
Db 1697 HYVNH-----KRNEYSNNIESKFNYY-----DDINKK---EYEDENDIYFTTHSS 1742
QY 512 RGSDELTAKTNTIVGDNVAMIGYVSD 538
Db 1743 QGNNDL-SNDNYLSSEELNTDEYDD 1768

RESULT 13

US-08-169-927-2
; Sequence 2, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Mei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David
; REGISTRATION NUMBER: 24,743
; REFERENCE/DOCKET NUMBER: 75,976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-1022
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1612 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-169-927-2

Query Match 4.1%; Score 172; DB 1; Length 1612;
Best Local Similarity 19.5%; Pred. No. 0.00018;

Matches 200; Conservative 117; Mismatches 366; Indels 342; Gaps 47;

```
QY 4 VKNWAVATLVASILMGVVTAAHQDVESNNYNGVAEVTNER-----QANGQIG 53
Db 245 VSKNGA-AIEFNVGTGAG-----NLKGIIELTAAVAGKLSLGLGAANAVIG 291
QY 54 VD-----GKIIS---ANSNTSG-----STNOESSAT-----NNTENAV 84
Db 292 TDNGAGRAAGFIVSDNGNAATISQVYAKNWIQSANAGGOVTEHIVDVGLGGTTNPK 351
QY 85 VNESKNTNTENAVVNEKNKNTNT-----ENAVVNEKNKNTNTE-----123
Db 352 TADSK-----VILTENSFGSTNFGNLDQIVVPDPTKILKNFGIDGVKNNGNTAGVIT 404
QY 124 -NDNSQLKLTNNEQPSAQTAN-----144
Db 405 FNANGALVSASTDPNIATNTINAEAGVGVELSGIHIAELRLGNGSGIFELADGTWIN 464
QY 145 -----LKKLNPOAAKAVQ--NAKIDAGSLTDDQIN--ELKNKFNKSAEKGAKLTFKDL 194
Db 465 GPVQNQALMNNNALAAGSIQDLGSAITITGDIINGGVNAAHQHITLNDASKILALDGANI 524
QY 195 EG--IGNAIVKODPOYAIPYNAKEIKNMPATYTVDAQTGKMAHLDVWDSNPVQD--PVTG 251
Db 525 IGANVGGAIHFQANGGTIKLTNTQ--NNIVVNFOLDITTDKTGVVDASSLTNNQTLTING 582
QY 252 YVSNYMGQVLVIAMKIPNSPT-----GD-----NHIYLLYKYGDNDFSHWRNA 296
Db 583 SIGTVANTKTLAQNLNGSSKTLNAGDVAINELVIENNGSVQLMHHNTYLITKTINAANO 642
QY 297 GSI-----FQTKETNV-----FQEWSSAIV 317
Db 643 GQIIIVADPLNTNTLADGTLNLSAENPLSTIHFATKAANADSILNVCKGVNLYANNITT 702
QY 318 NDDGTIOLFFTSNDTS-----DYKLNDRQLATATLNLNVDDNGVSGVSDNYQVL 367
Db 703 NDANVGLSHFRSGGTSIVSGTVGQOQGHKLNLL-----DNCTTVK-----744
QY 368 FEGDGFHYQVTEQFANGKREDDDYCLDRPHVQLENG--DRYLVFEANTGTEDYQSDQ 425
Db 745 FLGD-----TTF-----NGGTK-----IEGKSIQISNNYTTDHVESADNTGTLEFVNTDP 790
QY 426 I-----YNWANYG-----GDDAFNIKSSFKLLN-----NKKDRELALANG-----461
Db 791 ITVTLNKGAVFGLVKQVILISPGNIYFNEIGNVGIHVHGIANSISFENASLGTSFLPLS 850
QY 462 --ALGILKLTNNQSKPKVEEYVSLVTLMACDEVXXK---LGDKYILFSVTRVSRGSDR 516
Db 851 GTPDLVLTIKSTVGNGTVDNENAPIV-VVSGIDSMINNGIIGDKK---NIIALSGLSDN 906
QY 517 ELTAKDNTIVGDNVAMIGYVSDSLMGKVKPLNNSG--VVLTSASVAN-----561
Db 907 SITVNANTLYS-----GIRTTKNNOGTVTLSGMPNPNPTIYGLGLENGSP 952
QY 562 -WRPATY-----SYAVPVAGHPQVLITSYMSNKNDF-----ASGEGNVATWAP 604
Db 953 KKKQVTTFTDYNLGSIIANNVTINDVLTITGGIAGTDFDAKITLGSVGNANV-----1007
QY 605 SFLVQIIPDDTTTLARATNOGD-----WVDDSSRNDNMGLVKLKEGAANSALPGE- 656
Db 1008 RFVDSFSDPRSMIVATQANKGTVYTLGNALVSNIGSLDTPVASVRFTGNDSGAGLQGN 1067
QY 657 WGRKPDW-----SLINSGPLGLKPHQVPQKIDQDPQPSQNTKNTVTPNGDKPAGKA 711
Db 1068 YSQNIDGTVNLTILNSNVLIG--GGTTAINGEID-----LTTNNLIIFANGSTWGDN 1118
QY 712 TPDNTNDPSAQPQSQ---NTNIDPSAQSQNTKNTVTPNGKQKNTDAKQLPOTGNK- 767
Db 1119 TSISTLLNVSSGNTIGQVVIADQAVNATTGTGTTIKIQDNANANESGQATYLTQQGARF 1178
QY 768 SGLAG 772
Db 1179 NGTLG 1183
```

RESULT 14

```
US-09-336-447A-9
; Sequence 9, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-9
```

Query Match 4.1%; Score 170.5; DB 4; Length 941;
Best Local Similarity 20.7%; Pred. No. 9.8e-05;
Matches 190; Conservative 108; Mismatches 380; Indels 239; Gaps 41;

```
QY 4 VKNWAVATLVASILMGVVTAAHQDVESNNYNGVAEVTNERQANGQIGVDG-----KI 58
Db 111 IKNSTV-----GGFTNEA-----MGEYSTVAG-CANNOAKNGYVTVGGNGNKA 155
QY 59 ISANSTTSGSTNQ-----ESSATNTENAVVNESKNTNTENAVVNEKNNTN 106
Db 156 IGNSTVSGSNNOAKGEHSTTAGGKNNOATGNGSPAAGVENKADAN--NAVALGNKNTI 213
QY 107 NTENAVVNEKNNT-----NNTENDNSQ-----LKLITNNEQPSAATQANKLKN-- 149
Db 214 EGTNSVAIGSNNTVKTGKENVILGSENTENAQSGSVLLGNNTAGKAAATVNNAEVNG 273
QY 150 -----PQAKAVQN--AKIDAGSLTDDQINELKNKINFSK---SAEKGAKLTFKDLBG 199
Db 274 TIENFAGASKANANNITGTVSGS---ENNERQIVWAGAGISATSTSDAYNGSQLHALAK 329
QY 200 AIVKODPOYAIPYNAKEIKNM-PATYTVDAOTGKMA-----HLDVWDSMPVQDPVTGY 252
Db 330 AVAK-----NKSIDIKLNGKVKELDKVGVLSRDINSLHDDVDAN--QDSIAKN 376
QY 253 VSNYMGYQVLVIAM-----GIPNSPTGDNHIYLLYKNYKGNDFSHWRNAGSIFG-----301
Db 377 KADIKLNGKVKELDKVGVLSRDIGS-----LHDDVDANQDSIAKNKADIKLNGKVEKE 431
QY 302 -TKETNVFQWGSALVNDGDTIQLFFTSNDTSYKLNDRQLATATLNLN--VDDNGVS 357
Db 432 LDKVGVLSRDIGS--LHDDVDATNQADIAKNOAQNTDQDLAAVNLQDQVQAQKTEIDALNKA 489
QY 358 IKSVDNYQVLFEGDGFHYQVTEQFANG-----KRENDYDCLDRPHVQV 401
Db 490 DNNINNIYLAQQOQDQHSDDIKTKNNVEEGLDLISGRLLDQKADLTOKTKLKN-----545
QY 402 LENG-----DRYLVFEANTGTEDYQSDQI---YNWANYGDDDAFNK 441
Db 546 VEEGLDLISGRLLDQKADIAKNOAQNTDQDLAAVNLQDQVQAQKTEIDALNKA 605
QY 442 SS-----FKLLNNKKRELACLANGALGILKLTNNQSKPKVEEYVSLVTLMAC 491
Db 606 SSANTDRITATABELGAENKKDAQIA-----KAQANENKDKGI-----AK 643
QY 492 DEVXXKLGDY-----YLFVSVTRVSRGSDRELTAKNITVGDNVAMIG-----534
Db 644 NOADIQLDHDKKITNLGILHSWVARVGNNTQGVATNKAQADIAKNOADIANNIKNIYELAQ 703
```


Db	130	IYLFYQKAGNTSLSGWKKAGRVFKDNDKLLANDPILKNQSEEWSSGATLTSDGQVRLFTY	189
Qy	329	SNDTSDYKLN---DORLATATLNL-VDDNGVSIKSDYDNVQVLFE-GDGFYQTYEQFAN	383
Db	190	SRQYPDPNNLYGKQTLSTAQINVQPDOKTLKIDGVEDLKSIYDGDGKIYQVQO-SV	248
Qy	384	GKRENDYCLRDHPHVOLENGDRYLVFEANTGTE-DYQSDDOIYWNANYGGDADPNIKS	442
Db	249	GVDMDN---HTFRDPHYVE-DOGHKYIIFEANTGTGTGYQGEDSIQNPVYVGGNKKFTEE	305
Qy	443	SFKLLNNKDLRELAGLANGALGILKLTNNOSKPKVEEYSPLYSTLMACDEV---XXKL	498
Db	306	QONLLQSPK-KGAELANGALGIVELNDDTY---LKNVMPPLASNLVTDIEFRANVFKM	361
Qy	499	GDKYLFVSITRVSGSDRELTKADNTIVGDNVAMIGYVSDLSMGKYKPLNNSGVILTASV	558
Db	362	NGLWYLFSTRGSK-----VTVDAGDDDDIYMLGYVSTLTGPYKPLNGTGLVLHQDL	414
Qy	559	PANWRATYSYVAVPVAGHPDQVLITSYMSKNKDFASGEGNYATWASFLYQINPDOTTIV	618
Db	415	DRDDVTWYTAHFAIP-OGKGNVNVVVSMTNRGLF---PDHKSTFAPSLNINIKGSTSV	471
Qy	619	LARATNGDQVWDDSSRNDNMLGVLKEGAANSALPGEWCK	659
Db	472	KNGLEOGQITVDPT---NDK-----PNEYCK	498
RESULT	3		
O82854			
ID	O82854	PRELIMINARY;	PRT: 487 AA.

01-NOV-1998 (TREMUREL. 08, Created)
DT 01-NOV-1998 (TREMUREL. 08, Last sequence update)
DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
DE BETA-FRUCTOFURANOSIDASE.
OS Bacillus sp.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1409;
[1] SEQUENCE FROM N.A.
RC STRAIN-V230;
RA Tsusaki K.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[2] SEQUENCE FROM N.A.
RC STRAIN-V230;
RA Kurimoto M.;
RT "Cloning and sequencing of beta-fructofuranosidase gene from Bacillus
sp. V230.";
RL EMBL: AB010272; BAA32083.1; -
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pram; PF02455; Glyco_hydro_68; 1.
SQ SEQUENCE 487 AA; 53412 MW; 7F52A47921824AE3 CRC64;

Best Local Similarity 35.3%; Pred. NO. 2.4e-25;
Matches 168; Conservative 84; Mismatches 141; Indels 83; Gaps 19;
QV 177 INFKSKEGAKLTFKDLGEG-----NAIVKQD----- 205

Db	1	MNFRLAKLAAAVTFRYAILVAGDPHIFIAQNMNSGGYKEDYGFPAHITRADMLKIPCCQN	60
QY	206	-PQYAIPIYSNAKEIKNPPATYTVDAQTKMAHLVDVDSWPVDPVTGVSYNMGYOLVIA	264
Db	61	SPOPKVPOFNASAIKNTDSAKGYD-KSGNLTLDVDSWPQLN-ADGTAANYHGYHTVSA	118
QY	265	MWGPNSPTGDNHLYLLYLNKYGNDFSHWRNAGSIFQTKETNV-----PQWSSGSA	315
Db	119	LADGPKN-SDDTPHLHLYQKVGDPSTIDSQWNAKRGVDEDMKFPNDPILKYOTQWSSGSA	177

Qy	316	IVNDGDTQIQLFTSNDTSYKLN-----DQRIATATLNLNVDDNG--VSIKSVSDNYQ 365
Db	178	TLTKDQGVRLRYFT-----DTSGNPDGCTGAGNQIISTAQVNLSPQDAATLKV DGVGSDHK 232
Qy	366	VLFE-GDGFHQYTYEQFAN-GKORENDYCLRDPHVHVLGNGDRYLVFEANTCTED-QYS 422
Db	233	SVFDGGDGTYYNIQIQFIDEGKKWISGDNHTLRDPHYVE-DKGHKYLVFEANTGTTDGYOG 291
Qy	423	DDQIYNWANYGGDDAFWNKISFSFKLLANKKDRELAGLANGALGILKLTNNQSKPKVEEVS 482
Db	292	DQSFNNKAYYGGSDVPFQNEKNKLLQSPK-KQIASLANGAIGIVELADDYT---VKSVMK 347
Qy	483	PLVSTLMACEV---XXKLGDKYVLFSTVRSRSGSDRELTAKDNTVIGDNTVAMIGYVSD 538
Db	348	PLVASNTVADEVANETFRMNNKXWYLFSTSRGSKMTSDGINDK-----VYMLPGGD 400
Qy	539	SLMGKYPKLNNSGVVLTSASVPANNRATSYAYAVPVAGHPD--QVLITSYVMSNKDF 592
Db	401	SLNGPHNPINETGLVNLNMNLDPADLTHTYSHCGIP---HPEGNNVLTYSYMTNRGF 453

RESULT 4

Q97179

PRELIMINARY; PRT; 428 AA.

Q97179

Q97179; 01-OCT-2001 (TrEMBLrel. 18, Created)

01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

LEVANSUCRASE.

CAC1174.

OS Clostridium acetobutylicum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1488;

XP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

RC MEDLINE=21359325; PubMed=11466286;

RA Neolling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,

RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,

RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,

RA Bennett G.N., Koonin E.V., Smith D.R.

RT "Genome sequence and comparative analysis of the solvent-producing

RT bacterium Clostridium acetobutylicum.";

RL J. Bacteriol. 183:4823-4838(2001).

DR EMBL: AE007686; AAK79739.1; -

DR InterPro: IPR003469; Glyco_hydro_68.

DR Pfam: PF02435; Glyco_hydro_68; 1.

Complete proteome.

SW SEQUENCE 428 AA; 48395 MW; 9670B154B178E23E CRC64;

[illegible]

Qy	467	KLTNQSKPKVEEYVSLVSTLMACDFVXXKLGDKYILFSVTR---VSRGSDRELTAQDN	523
		: : : : : : : :	
Db	285	DLTRFKLLPPLLEAVG--VNOQLERPOIVMK-KNKYILFTISHKFTYAQG---LNGVD-	336
		: : : : : : : :	
Qy	524	TIVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANRWRTATSYAVVPVAGHPDQVLI	583
		: : : : : : : :	
Db	337	-----GLYGFCGNSLSRNYKPLNGNGLVITN--PTNDPQIYSWYL--VSGHDVLSFI	395
		: : : : : : : :	
Qy	584	TSYMSNKDFASGEGNYATWAPSLVQI	610
		: : : : : : : :	
Db	386	NEYHFNGQLRYG-----GTFAPTQLISL	408
		: : : : : : : :	

RESULT 5

Q9EVD6	PRELIMINARY;	PRT;	630 AA.
ID	Q9EVD6		
AC	Q9EVD6;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	FRUCTOSYLTRANSFERASE.		
GN	FTF.		
OS	Actinomyces naeslundii.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.		
OX	NCBI_TaxId:1655;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-ATCC 12104, WVU45;		
RA	Bergey L.J., Morou-Bernudez E., Burne R.A.;		
RT	"Characterization of the fructosyltransferase gene of Actinomyces naeslundii.";		
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF228582; AGO9737.1; -		
DR	Interpro: IPR003469; Glyco_hydro_68.		
DR	Pfam: PF02435; Glyco_hydro_68; 1.		
KW	Transferase.		
SQ	SEQUENCE 630 AA; 68256 MW; CDCA04184DD6D7D2 CRC64;		

Query Match	5.9%;	Score	245.5;	DB 2;	Length	630;			
Best Local Similarity	22.7%;	Pred. No.	0.00013;						
Matches	150;	Conservative	91;	Mismatches	285;	Indels	135;	Gaps	

Qy	75	SATNNTENAVVNSKNTNTNENAVVNENKNTNTNENAVVNENKNTNTNTENDNSQLKTN	134
Db	26	AGTGTALADETSPPTSSSATAEASPEAGADOKNDSPADQCATQTPTDQCAQADQ	85
Qy	135	EQP--SAATQAN-LKLNQAAKAVONAKIDAGSLTDQINELKNINFSKSAEKGAKLTF	191
Db	86	QAAGGAAQAQANGFKAQADPNCPANAKHTGAAGH-----VEENYAKWTPADAMQTORVF	139
Qy	192	KLEBIGNATVKQDOQVAIPYSNAKEIKN-MPATYTVDAQTGKMAHLVDWDSMPVDPVT	250
Db	140	QP-QGAVRAPTRLPEQLTMB-----EISNGEPAT-----SEDVWVWDVTWLTDEAA	184
Qy	251	GYVSNYMGYQLVTAMGIPNS--PTGDNHIY-----LLNKKYGDN----DFSHWRNAGSIF	300
Db	185	HOIS-YNGWEIAFSLVADRHAGYTFDDRTHARLGLGFYRKAGTQTSSADGAGSSNGGIWY	243
Qy	301	G-----TKETNVF--QEHSGSAIVNDGDTQIOLFTS-----NDTSQY-----	335
Db	244	GGHVFPDASARSRTVRHAQTEWSSGARLMEGNKKIRMFYTSVPSTAPPRTGTTTHQVTH	303
Qy	336	---KLNDORLATATLNLVNDNGVSIKSDVDNYQVLQEGDGFHQYTYEQFANGCKDRENDY	392
Db	304	ATSKPYDPRIQSEGRIYATKNGVWLTGFTQHQLLVPDQKYYQTRQ-----NPGV	355
Qy	393	CLRDPHVVQLENG-----DRYLVEAMTGTEDYQSDQOYINWANYGGDDAFNIKSSPKLN	448
Db	356	NFRDPPTFRQONNPSDPTVWVEGNSA-----FVREQYYDAAKAGQNTLTATCEEDLG	412
Qy	449	NKK---DRELAGLAGALGILKLTN-----NOSKPKVE-EVYSPLYSTLWACDDEVXXK---	497

```

Db 413 YEKDPRKAEVAVNQGGYQLANGLAKRANKAMTEWEYLPPLLSGNCVNDQTERPQI 472
QY 498 --LGDKYLFVSVTRSGDSRELTAKONTIVGDNVAMIGYVSDSLMGKYKPLN-NSGWL 554
Db 473 YFODGKYLFETI-----SHRETYA--DGLQGPB-GYVGFVGDGLRSDYKPLNQNTGIAL 523
QY 555 TASYPANWRTA-----TYSYAVPVVACHPQOVLITYSNKDKFASGEGNY 599
Db 524 GNPINLNFCKVPSPDNOSPYTFQSYHYVMP--GGLVESEFDSISGGKNGKNPVRG-- 579
QY 600 ATWAPSLVQINPDPTTIVLARATNOGDWDDSSRRNDMLGVLEKGAANSALPGEWGK 659
Db 580 GSLSPVTKLNIS-GDTTSV-----DRTYCTNGLGGFADIPADRATNGGATR 625
QY 660 P 660
Db 626 P 626

RESULT 6
ID Q93FU9 PRELIMINARY; PRT; 424 AA.
AC Q93FU9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE LEVANSUCRASE.
GN LSCA.
OS Pseudomonas aurantiaca.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=86192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S-4380;
RA Koh I., Jang E.K., Kim J.S., Song K.S., Kim C.H., Rhee S.K.;
RT "Screening, molecular cloning and characterization of a novel
RT levansucrase, and its expression in Escherichia coli.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF306513; AAL09386.1; --
SQ SEQUENCE 424 AA; 47009 MW; 9C1B4936A985162E CRC64;

Query Match 5.7%; Score 238.5; DB 2; Length 424;
Best Local Similarity 22.9%; Pred. No. 0.00018;
Matches 95; Conservative 76; Mismatches 152; Indels 91; Gaps 21;

QY 239 VVDSWPVQDPVTGYVSNMGYQLVIAMWG--IPNSPT-----GDNH-----I 278
Db 51 INDWPLRD--LDGNTVSDGVSVIFLTADRHNPNDPEYIDENGNYDITRDWDRHGRAMK 109
QY 279 YLLNYKGNDFSHWRNAGSIFGKTETNVFQWGSAGI--VNDGDTIQLPFT--SNDTSDY 335
Db 110 YVWFSRTGKD---WKLGRVMAEGVSPAREWAGTPIILLNQGVDLYTAVTGCATIV 165
QY 336 KLNDQRLATATLNLVNDNGSVIKSDVNVQVLFEGDGFHYQYFQFANGKRENDYCLR 395
Db 166 KVRGRVVT-----EHGVSMVGFEKVKPLFEADGKMYQTEAQ--NPFVWGRDPWPF 215
QY 396 DPHVQLENGDRYLVFEANTGTEDYQSDDIYNWANYGGDDAFNIKSFLLNNKKDR 455
Db 216 DPN-----DGKLYMFEQNVAGE--RGSHKV-----GKAEIGDPPGVEDVGNR----- 258
QY 456 AGLANGALGILKLTNNQSKPKVEEYVSPVLSTLMACDEVXXK----LGDKYILFSVTRVS 511
Db 259 --FQACVGIARVADGDD--WEMLPPLLTAVGVNDQTERPHFVFQDGKYLFTISHT- 313
QY 512 RGSDELTAKTNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYSYA 571
Db 314 -----FTYADGVTPGDGV--YGFVADSLFGYVPLNGSGLVL--GNPSSQPFQTYSHCV 363
QY 572 VPVAGHPDQVLITSYM-----SNKDFASGEGNYATWAPSLVQINPDITTV 618

```

```

Db 364 MP-----NGLVTSFIDSVPDESQTQIRIG-----GTEAPVVEIKIKGQTFVV 407

RESULT 7
ID Q93TMI PRELIMINARY; PRT; 431 AA.
AC Q93TMI;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE LEVANSUCRASE LSCC (EC 2.4.1.10).
GN LSCC.
OS Pseudomonas syringae (pv. glycinea).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PG4180;
RA Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT levansucrase genes from the plant pathogens Pseudomonas syringae pv.
RT glycinea and P. syringae pv. phaseolicola.";
RL Appl. Environ. Microbiol. 64:3180-3187(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PG4180;
RX MEDLINE=21242726; PubMed=11344135;
RA Li H., Ullrich M.S.;
RT "Characterization and Mutational Analysis of Three Allelic lsc Genes
RT Encoding Levansucrase in Pseudomonas syringae.";
RL J. Bacteriol. 183:3282-3292(2001).
DR EMBL; AF346402; AAK4952.1; --
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 431 AA; 47619 MW; 0C71D0F133071AA5 CRC64;

Query Match 5.7%; Score 237.5; DB 2; Length 431;
Best Local Similarity 23.2%; Pred. No. 0.0002;
Matches 95; Conservative 66; Mismatches 154; Indels 95; Gaps 18;

QY 239 VVDSWPVQDPVTGYVSNMGYQLVIAMWGIPNSPTGDNH-----IYLLN----- 283
Db 60 INDWPLRE--LDGTVVSVNGWSVLT-----TADRHPPDPQYLDANGRYDIKRDWED 111
QY 284 KYGDNDFSHWRN-----AGSIFGKTETNVFQWGSAGI--VNDGDTIQLPFTSNDTSD 334
Db 112 RHGRARMCYWYSRTGKDWIFGGRVMAEGVSPTRWAGTPIILLNDKGDIDLYTCTVTPG- 170
QY 335 YKLNQRLATATLNLVNDNGSVIKSDVNVQVLFEGDGFHYQYFQFANGKRENDYCLR 394
Db 171 -----AAVAKVRGRIVTSQGVLEKDFQVKKLFEADGTYTQT-----EAQNSWNF 217
QY 395 RDHP--VVQLENGDRYLVFEANTGTEDYQSDDIYNWANYGGDDAFNIKSFLLNNKKDR 453
Db 218 RDPSPFIQNDGKLYVFEQNV-----AGERSHTVGAELGPPVPGH 260
QY 454 ELAANGALGILKLTNNQSKPKVE--EYVSPVLSTLMACDEVXXK----LGDKYILFSVT 508
Db 261 EDVGGARFQVGCIGLAVAKDLSGEWEILPPLVTAAGVNDQTERPHYVQDGKYLFTIS 320
QY 509 RYVSRGSDRELTAKTNTIVGDNVAMIGYVSDSLMGKYKPLNNSGVVLTASVPANWRTATYS 568
Db 321 -----HKFTYADGIGTPDGV--YGFVGEHLFGYRPMNASGLVL--GNPPEQPFQTY 369
QY 569 YVAVPVAGHPDQVLITSYMSN-----KDFASGEGNYATWAPSLVQINPD 613
Db 370 HCVM-----NGLVTSFIDSVPTEGEDYRIG-----GTEAPVIRILLKGD 409

RESULT 8

```

Q9AEX9
ID Q9AEX9 PRELIMINARY; PRT; 415 AA.
AC Q9AEX9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE LEVANSUCRASE.
OS Rahnella aquatilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Rahnella.
OX NCBI_TaxID=34038;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim H., Cha J.;
RT "Cloning and characterization of levansucrase";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY027657; RAK14794.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
SQ SEQUENCE 415 AA; 45958 MW; F09F5A1BFB690C5E CRC64;

Query Match 5.7%; Score 235.5; DB 2; Length 415;
Best Local Similarity 23.2%; Pred. No. 0.00024;
Matches 92; Conservative 69; Mismatches 151; Indels 85; Gaps 18;

QY 239 VDSWSPVQDPVTGYVSYNGYQVLIAMGIPNSPTGDNHIYL-----LYNKYGD 287
DB 44 INDAMPLRS-LDGTVVSDGWSVIFTLTAQRNN--NNSEYLDADGNYDITSDWNNRHGR 99
QY 288 NDFSHRN-----AGSIFGKTKENYQFWSGSAI-VNDGTTQLFTSNDTSYKLN 338
DB 100 AKICYWYSTGKDWTFGRVMAEGSPPTREWAGTPILLNEDGEVDLYTCTVTPG----- 154
QY 339 DQRLATATLNLVNDGWSIKSDYQVLFEGDGHYQTYEQFANGKRENDYDCLRDHP 398
DB 155 -ATIAKVRGKVLTSBEGVTLAGFNEVKSLFSADGVYQTESQNTYWNFPDPFFI--DPH 211
QY 399 VVQLENGDRYLVEANTGTEDYQSDQIYNWANYGGDDAFNTKSSFKLLNNKKRELALG 458
DB 212 -----DGKLYMYFEGNVAGE-----RGSHVIGKQEMGTLPDHRDVGARY 252
QY 459 ANGALGIL---KLTNOSKPKVEEYVPLVSTLMACDEVXXK---LGDKYLYLFSVTRVS 511
DB 253 QAGCIGMAVAKOLSGDE-----WEILPELVTAAGVNDQTERPHFVQDGKYYLFTISHKF 307
QY 512 RGSRELTAKNWTIVGDNVAMTGYVSDLMGKYKPLNNSGVVLTASVPANWRTATYSYA 571
DB 308 TYAE-GLTGPD-----GVYGFSLDRLTGTPSPMNGSGLVL--GNPPSQPFQYSHCV 356
QY 572 VPVAGHPDQVLITSYNSKNDFASGEGNY---ATWAPS 605
DB 357 MP-----NGLVTSFDN--VPTDGNRYIGTGTEAPT 385

RESULT 9

Q9AISO
ID Q9AISO PRELIMINARY; PRT; 1395 AA.
AC Q9AISO;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FIBRONECTIN BINDING AUTOLYSIN.
GN ATLC.
OS Staphylococcus caprae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=29380;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101014; PubMed=11159959;
RA Allignet J., Aubert S., Dyke K.G.H., El Solh N.;
RT "Staphylococcus caprae Strains Carry Determinants Known To Be Involved

RT in Pathogenicity: a Gene Encoding an Autolysin-Binding Fibronectin and
RT the ica Operon Involved in Biofilm Formation.;
RL Infect. Immun. 69:712-718(2001).
DR EMBL; AF244123; RAK17065.1;
DR InterPro; IPR002502; Amidase_2.
DR InterPro; IPR002901; Amidase_4.
DR Pfam; PF01510; Amidase_2; 1.
DR SMART; SM00047; LY22; 1.
SQ SEQUENCE 1395 AA; 153796 MW; 1DEC54FA60035F2E CRC64;

Query Match 5.6%; Score 233; DB 2; Length 1395;
Best Local Similarity 20.7%; Pred. No. 0.0014;
Matches 175; Conservative 105; Mismatches 299; Indels 266; Gaps 40;

QY 28 ADQVESNNYN-----GVAENVERQA-----NQIGVDGKIISANSNTSGSTNQES 74
DB 104 SQOEQSSNNQPTDSKETNTSSVNTYQDSTKDIKNEISKDAVTSNEDNETSKVDNTNKEV 163
QY 75 SATNNTENAV--VNESKNTNNTENAVVNNKNTNNTENAVVNNKNTNN-----T 122
DB 164 SETEQOESTVENYKDESNNNLQASSNQTNNLNDQEKQEATSNQASNKNTSQOATDST 223
QY 123 ENDNSOLKLTNNE-----OPSAATQANLKLNPAAKAVQNAKIDAGSLTDDQI 171
DB 224 EKSNNKVSQTSNSGYNFDDDDVDTTNAQSKNTKADQPOVASLSAQNNVNTKESNPTI 283
QY 172 -NELNKNFSKSAEKGAKL-----TFKDLGIGNAIVKQDPOYAIPIYS 213
DB 284 KSTNSTNTSKGATQNTKVSNTQANNNTTESNNKYSTF-----SSVAKPRMYAV--- 333
QY 214 NAKETKNPATYTVDAQTKMAHLVDWSPVQDPVTGYV--SNYMGYQLVITAMMGIPNS 271
DB 334 NKTTSLP-KYTPQ-----VKSSINDYIRKKNY-----KA 363
QY 272 PTGDNHIYLLYKNYGNDFSHRNAGSIFGTKETNVFQEWGSAIVNDDGTIQLFTSND 331
DB 364 PQTEENYTSYFPKYG-----YRNG---VGRPE-----GIVVHD-----TAND 397
QY 332 TS-----DYKLNDRATATLNLVNDGWSIKSDYQVLFEGDGHYQTYEQFANGK 385
DB 398 NSTIDGEINMYKNYQ--SAFVHAFVDGHRIVETAPDI--LSWGAG--PQGNREFINVE 451
QY 386 DRENDYDCLRDHPVQLENGDRYLVEANTGTEDY-----QSDQIYNWANY----- 432
DB 452 IVTHDY-----DSFARSMNMYADY---AATQLYYGLKPSAENDGQGTWTHYALSRYL 504
QY 433 GGDADFNIKSSFKLLNNKKDRELALGALGILKLTNNQSKPKVEEYVPLVSTLMACD 492
DB 505 GGTDHDPHQYFR-----SHNYS-----YAEYDLIY--- 531
QY 493 EVXXKLGDKYLL-----FSVTRVSRGSDRELTAKNQNTIVGDNVAMIGYVSDSL 540
DB 532 -----EKYLKTQGVAPWGTSTSPSKPSKPSDSSNNKLTVSANSVAQI----- 577
QY 541 MKKYPLNNSGVVLTASVPANWRT---ATYSYVAVPVAGHPDQVLITSYNSKNKDFASGE 596
DB 578 -----KP-SNSGLYTTVYDEKGHSTDAQKTLSTVTSATLGNKKEVLVEDYNTGKKYV--- 629
QY 597 GNYATWAPSLVQINPDDTTTLARATNQGVDWDDSSRDNHMLGVLKEGANSALPGE 656
DB 630 -----W-----VKQGDVYV--NTAKSPVKVQNTVNVKAGSTLYTVP 663
QY 657 WGPVDMWSLINRSPGLKLPHPV--QPKIDQDPQSPSGONTKN-----VTPGNGDKP 707
DB 664 WGTSP--QVASKVSGSGNQTFKATQQQQIDKAIYLYGVNGKSGWISKYLLTTPSSNTK 721
QY 708 ACKATPDNTNIDPSAQPQSGQNTNIDPSAQSGQNTKNTVTPGNEKQGNKNTDAKQLPQTGNK 767
DB 722 FSKPSTDNSSNNKLTVSA-NSGV---AQIAKNNGVTTVYDKKGTQDQVQRTLSYTK 777
QY 768 SCLAG 772

Db 778 SATLG 782

```
RESULT 10
Q9KWX6 PRELIMINARY; PRT; 1171 AA.
ID Q9KWX6
AC Q9KWX6
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE BONE SIALOPROTEIN-BINDING PROTEIN.
GN BOP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=024;
RX MEDLINE=20115096; PubMed=10642520;
RA Tung H.S., Guss B., Hellman U., Persson L., Rubin K., Ryden C.;
RT "A bone sialoprotein-binding protein from Staphylococcus aureus: a
RL member of the staphylococcal Sdr family J.";
RL Biochem. J. 345:611-619(2000).
DR EMBL: Y18653; CAB75732.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 1171 AA; 127123 MW; C5BC812F9DA5A884 CRC64;

Query Match 5.5%; Score 230; DB 2; Length 1171;
Best Local Similarity 22.0%; Pred. No. 0.0016;
Matches 167; Conservative 103; Mismatches 296; Indels 194; Gaps 35;

QY 12 TLVSASALMGVV-----TAHADQVE---SNNYNGVAEY-NTREQANG- 50
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 29 TVGTASILVGTTLFGLGNOEAKAENTSTENAKQDEASADNKEVYSETENNSTQKNDL 88
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 51 -----QIGVDGKIISANSNTSGSTNOESATNTENAVVNSKNTNTENAVVNSKN 104
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 89 TNPKEKTNDSHQEAKETPTSTQQQNNATSTETKPNIEK-----ENKPSDKT 143
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 105 TNNTENAVVNSKNTNTENDNSQLKLTNNQPSAAT---QANLKLNPQAQAKAVQNAKI 161
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 144 ATEDTSVILEEKAPNTNDVTTPKSTSEIQTPPTPQESTNIENSQPTP-----SKV 199
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 162 DAGSLTDQQLNELNKNFSKSAEKAFLFKDL-EGIGNAIVKQDPQYATPYSNAKEIKN 220
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 200 D--NQVTDITNPKPEVNSKEELKNPEKLELVRNDSNTDRSTKPVATAPTSPAPKRVN 257
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 221 MPATYTVDAQTGKMAHLVDWSDWPQDPV-----TGYVSNMGYQLVTA----- 264
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 258 AKIRFAV-AQPAAVASNNVNDLITVTKMITEGIKDDGVIAQHDGEHIIYTSDFKIDNAV 316
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 265 -----MMGIPNSPTGD-----NHLYLY----- 282
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 317 KAGDTMTVKYDKETIPSDITDDFTPDITDPSGEVIAKGTFDLNTKTIYKFTDYVDRE 376
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 283 -----NKYGD-NDFSHWRNAGSIFGT--KET--NVQEWSSGSAIVNDGCTIQLFFTS 329
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 377 NVAKLELSYIDKKEVPNETNLNLFATADKTSKRVKEYQ-KPIVKDESNIQSFISH 435
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 330 NDTSDYKLNLDQRLATLNLNLDNGVSIKSDYNDYQVLFEGDGFHYQYEQFANGKDRDN 389
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 436 LDYTKHEV-EQTIYVPLKLNKANTNTIKS-----GGVADN 471
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 390 DDYCLDRPHVQVLENGDRYLVFANTGTEDYQSDDDOYIWNWYGGDAPFNKSSFKLLNN 449
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 472 GDYDTGDSGI--IDSNTETKVKYKVASGQLPQS--NKIYDYSQY--EDVTSNVT-----INK 523
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 450 KKDRELAGLANGALGILKLTNNQSKPEVEVYSPVSTLMACDE-VXXKLGDYKYLFS-- 506
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 524 NYGTNMANIFGDISAYIV-----KVYSKYTPGAEDDLAVQOQGVMTTNTKINYSVA 577
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 507 -----VTRVSRGSDRELTAKDNTIYGDVNAVIGYV-SDSLMGKYKPLNNSGV 552
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 578 GYTILFQQLLTVVTVSVKPEEL-KIGDYVVEDVDKVGQGTDS---REKPMAN--V 631
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 553 VLTASVP-----ANMRTATYSYAVPVAGHPDQVLIT-----SYMNSKDFASGEENYAT 601
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 632 LVTLTYPDGTTSKVRTDANGHY--EFGCLKDGETYVTKFETPAGYLPPTKENGTTDGEKDS 689
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 602 WAPSLFQVQINPD-----DFTTVLARATNOGDVWDDSSRN 636
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 690 NGSSVTVKINGKDDMSLDTGFYKEPKYNLGDYVWEDTNKD 729
|: |||||: | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
Q9U0K4 PRELIMINARY; PRT; 622 AA.
ID Q9U0K4
AC Q9U0K4
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL 73.3 KDA PROTEIN.
GN MALP2 18.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,
RA Quail M., Barrell B.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035475; CAB62859.1;
KW Hypothetical protein.
SQ SEQUENCE 622 AA; 73277 MW; EA70E3C3A9F94925 CRC64;

Query Match 5.3%; Score 219.5; DB 5; Length 622;
Best Local Similarity 20.8%; Pred. No. 0.0024;
Matches 104; Conservative 94; Mismatches 194; Indels 107; Gaps 22;

QY 32 ESNYNGVAEVNTERQANGQIGVDGKIISANSNTSGSTNOESATNN---TENAVVNES 88
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 103 QKNMNN--EELSEQYN-----DCNNINKNNKIDSIINNNDINNENINNDNSINN 155
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 89 KNTNTENAVVNSKNTNTENAVVNSKNTNTENDNSQLKLTNNQPSAATQANLKL 148
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 156 DNINNDN--INNENINNNEN--INNNDINNDNINNDNINNDNINNDNINNSRPVVRKV 211
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 149 NPOAKA-VONAKIDAGSLTDDQINE-----LNKINFSKSAE-----KG 186
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 212 IKRVTRVANKKDETKNKLNDLNNQNSNDISCDVINKYKGTQDSDDEILNSIIQKN 271
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 187 AKLTFKDELGNAIVKQDPQYATPYS---NAKEIKNPATYTVDAQTGKMAHLVDWSDW 243
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 272 ELIMDPTTEGNKELTQKKKILIVRKINVSGRKKNVDDKNVYAE--KYNTSEIYNS- 328
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 244 PVQDPTGYVSNMGYQLVIAMMGIPNSPTGDNHIIYLLNKKYGDNDFSHWRNAGSIFGT 303
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 329 -----DENLF-----FGNKGKNNNL-SVEFEK 351
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 304 ETNVQEWSSGSAIVND-----DGTIQLFFTSNDTSDYKLNLDQRLATLNLNVDNGVS 357
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 352 KTHLISGONHTEILEDIENKSHDNVMDLKDKKKIKDYHMNN-----TSVKGLENEE----- 402
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 358 IKSV--DNYQVLFEGDGFHYQYEQFANGKDRNDYDCLDRPHVQVLENGDRYLVFEANT 415
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 403 IKSIEDRKHIOFSLDKILINNEMQRNTLNDNNKIKEIN-KELINEKLYLSKEME 461
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
QY 416 GTEDYQSDDDOYIWNWYGGDAPFN--KSSFLL-----NNKKDRELAGLANGALGILKL--T 469
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
DQ 462 LERKLESNEMYNLLN-----DKYNILENENKMLDRNEKELQDMNEEKLLKKIIEE 517
|: |||||: | | | | | | | | | | | | | | | | | | | | | |
```


QY 470 NMQSKPKVEEYSPVLSTL 488
DB 518 NNSLLKKKEQINEYINEI 536

RESULT 12

077151 ID 077151 PRELIMINARY; PRT; 930 AA.
AC 077151;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HC-23 PROTEIN.
GN HC-23.
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98398455; PubMed=9729442;
RA Schroeder A.A., Brown A.M., Abrahamson M.S.;
RT "Identification and cloning of a developmentally regulated
RT Cryptosporidium parvum gene by differential mRNA display PCR.";
RL Gene 216:327-334(1998).
DR EMBL; AF076438; AAC95010.1; -;
DR InterPro; IPR000571; Zf-CCCH.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00642; Zf-CCCH; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00356; Znf_C3H1; 1.
SQ SEQUENCE 930 AA; 98563 MW; 5AD78C87FB61F190 CRC64;

Query Match 5.1%; Score 212.5; DB 5; Length 930;
Best Local Similarity 20.8%; Pred. No. 0.0088;
Matches 174; Conservative 99; Mismatches 307; Indels 255; Gaps 39;

QY 21 GGVTAHADQVESNNYNGVAEVNTERQANGQIGVDGKIISANTSTSGSTN-----Q 72
DB 212 GGVLN---DGFNNHSHNSVALONN-----TVNHNHTSGSTNGSGFISDPWM 253
QY 73 ESSATNTNTENAVVNE-SKNTNTNTENAVVNEKNNTNTE-----NAVVNEKNNTNTENDN 126
DB 254 STSGYSPSNGFLHDFNSYDNNINSLNPNMNCNTEGISTDTNTNINHMQCNGANG- 312
QY 127 SQLKLTNNQPSAQAOLKLNFOAKAVONAKIDAGSLTDDQINELNKFNSKSA--E 184
DB 313 -----ILNNGQNSKIYNGAMNNLS-----ETRSIIETDKKYCDLINEGKNKDNLYQPGINV 363
QY 185 KGAKLTFKDLEGNAIVKQDPQVAIPYSNAKEIKNMPATYVDAQTGKMAHLVDWDSWP 244
DB 364 SFWNTAPNLISN-----NASDIYQNSIY-----WT 393
QY 245 VDPVTVGVSNMYGQYLVAMMGIPNSPTGDNHIIYLLYNKYGDNDFFSHWRN-AGSIFGK 303
DB 394 VQNPW---ACNSM-----NSFNQANNHGIIPSTT 420
QY 304 ETNV-FQEW-----SGSAIVNDGDTIQLFTSNDTSYKLNQO---RLATATLNLNV 351
DB 421 SVVDQFORFCGLGFDINSONQVTCDSQNNKNNMLISEHINNVTNNRNNRNSNGGATLSS 480
QY 352 DNGVSIKSDVNVQVLFGEGFHYOT---YEQFANGKO-----RENDYCL---RDPHVQV 401
DB 481 HONKPLDNDNNACTGEGGDISVESGVNAYINGSNLTENKLENRLQIGSNMNDHLIA 540
QY 402 L--ENGDRYLVEANTGDTYQSDQIYNNANY-----GGDDAFNIKSSEKLL 447
DB 541 INEDSGSNNI-----NSSINDTNTQLFNSSAPPLTSALIKQIPSSKSTKISVTSI 596
QY 448 NKKDRELAGLANCAGILKLTNNQSPKVEEYSPVLSTLMACDEVYX---KLGDKYYL 504
DB 597 NNNN-----GLNTGNSQNECLNSSNPKLSS-----ISTSGSNQNNNNNPNSTGSPKL 646

QY 505 FSVTRV-SRGSRE-----LTA---KONTI-----VGDNVAMIGYYS---DSLGMKY 544
DB 647 ISMVGVRNEGSIRNNLIGSALSANNSKQNTINSRPGVGVGNLNLSTGNYSDAVLHK- 705
QY 545 KPLNNSGVLTASVPANWRTATYSYAVPVAGHPDQVLTYSYMNKDFASGEGNYATWAP 604
DB 706 NDSNNGGVTTSS-----STGRSENGASVNSDMN----- 735
QY 605 SFLVQINPDDTTTTLARATNOGDWVDDSSRNDNMLGVLKEGAANSALPENGKCPVDWS 664
DB 736 -----ISNDSSNNLLSVNTD-----EDSS---TALGAHGEPPYNYKRALCRHWMRGYCW 781
QY 665 LIN-----RSPGLGLKPHQVPQIDQDQPSQNTKNV-----T 700
DB 782 EADCKFAHGEAEELTRDGLRHPTLSNGNSEVANQNGQSOSQASQYQSAQAQLPT 841
QY 701 PNGDKPAGKATPDNTNIDPSAQSCQNTNIDPSAQXGQNTKNVTPGNEKQGN 755
DB 842 VNNVTIPSGNLATATSTFTSTSGG---SSVSSSIQSKGLKQOS---GNGKNGSN 891

RESULT 13

Q95PH4 ID Q95PH4 PRELIMINARY; PRT; 1318 AA.
AC Q95PH4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HISTIDINE KINASE DHK (FRAGMENT).
GN DHK.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Anjard C., Loomis W.F.;
RT "The histidine kinases of Dictyostelium.";
RL (In) Inouye M., Dutta R. (eds.);
RL HISTIDINE KINASES IN SIGNAL TRANSDUCTION, pp.1-1, Academic press,
RL San Diego (2001).
DR EMBL; AF362374; AAK54093.1; -;
KW Kinase.
FT NON_TER 1
SQ SEQUENCE 1318 AA; 144944 MW; B1DA5CA0427066F6 CRC64;

Query Match 5.0%; Score 207.5; DB 5; Length 1318;
Best Local Similarity 19.5%; Pred. No. 0.024;
Matches 163; Conservative 116; Mismatches 306; Indels 249; Gaps 38;

QY 30 QVESNNYNGVAEVNTERQANGQIGVDGKIISANTSTSGSTNQESSATNTENAVVNESK 89
DB 229 ELKSNNDKIIE-NKENQENNNNNNN-----NNNNNNNNNNNNNNNTNINNTNN 282
QY 90 NTNNTENAVVNEKNKNTNTENAVVNEKN-----TNTENDNSOLKLTNNQPSAQAOL 145
DB 283 NDSNNNNNCINDLKNN 341
QY 146 KKLNPQAAKAVONAKIDAGSLTDDQINELNKFNSKAEKAGLTFKDLF-GIGNAIVKQ 204
DB 342 KKKKKNEFTVFSVIDSGSIDPYSTNLLFPFSLSS--YNNSTNTDGEGLGIAICKQ 400
QY 205 DPQVAIPYSNAKEIKNMPATYVDAQTGKMAHLVDWDSWPVQ-DPVTGVVSNMYGVQYVI 263
DB 401 -----LSNLANGEIKYETEMEKSVPELOV-----PMKCDSSISITSSN----- 439
QY 264 AMWGIPNPTG-DNHIYLLYNKYGDNDF--SHW-----RNAGSI-----FG 301
DB 440 -----NSTTNTNHYPRIMNNQSSKFFANSKWGEGKLTIVDDNPNIGKVIAHLEPPG 493
QY 302 TKEINVFOEWSSAIVNDGDTIQLFTSNDTSYKLN-----NDQR- 341

Db 494 EK - - - - -VFQRTTQSAL - - - - -YFERNCGDFNLILLDPLSLVIDEIKQMKODSSN 542
QY 342 - - - - -LATATL - - - - -NUNVDNGSVTSKVDNYQVLPEGGDFHQTVEQFANGKREND 390
Db 543 IIRNPLVIMCTAKLRKSLNVDNHYLYKPIKREQL - - - - -TVLSQLLPNTSTINP 593
QY 391 DYCLRDPHVQLENGDRVLYFEANTGTDYQSDDOQIYNWANYGGDDAFNIKSSKLLNKK 450
Db 594 IYSNQLN - - - - -NSGSSNGGGGGGGGGGGGGSSSNDIF - - - - -NK 634
QY 451 KORELAGLANGALGILKLTNNQSKPKVEEYSPVSLTLMACDEVXXKLGDKYLYFSVTRV 510
Db 635 TKLGSNISTG - - - - -IGNSGLINSNNIPTVTPSNIIPNLLSCQL - - - - - 677
QY 511 SRGSRELTAKNTVGNVAMIGVVSLSMGKYPKPLNNSGVVLPASVPANWRTATYSYI 570
Db 678 - - - - -LTSLNANIPQLTNDIGITNNN - - - - -INNSLMFTT - - - - - 709
QY 571 AVPVAGHPDOVLITSYM - - - - -SNKDFASGEGNYATWAPSLVQINPDPTTVILARATN 624
Db 710 - - - - -PNSTLSNGITGLDNNNSNDTGSIDNNSN - - - - -ISTNDNNNDYFIR - - - - -N 753
QY 625 QGDVMDSSRDNML - - - - -GVLKEGAANSALPGEWPKPVDWLSLNKSPGLGLKPHQP 679
Db 754 NGLPQNDMTYNNVYLNHQCVLPK - - - - -SUSVP - - - - -STPLSYNMLPTNLINAK - - - - - 802
QY 680 VQPKIDPOQPSGQNTKVTGNGDKKAGKATPNT - - - - -NIDSAQSPSGQNTNIDPSA 735
Db 803 - - - - -RSSLQPLNEN - - - - -SVLPTNTPTPLTSASQSLPMPGNDINSILPTQOOSQIDLOS 854
QY 736 QXSGQ - - - - -NTK - - - - -NVTPGNEKQKNTDAK - - - - -QLPOTGNKS 768
Db 855 QIELQLQSTIIRDRGDIILPDSTLEQITNLSSNSTISINPLPTNNNT 908
RESULT 14
Q97K42 PRELIMINARY; PRT; 2817 AA.
AC Q97K42
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE UNCHARACTERIZED PROTEIN, RELATED TO ENTEROTOXINS OF OTHER
DE CLOSTRIDIALES.
GN CAC1079
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AB007623; AAK79053.1;
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 119.
KW Complete proteome.
SQ SEQUENCE 2817 AA; 318224 MW; 1851D0D4FFBEE921 CRC64;

Query Match 4.9%; Score 203.5; DB 16; Length 2817;
Best Local Similarity 21.7%; Pred. No. 0.093;
Matches 143; Conservative 92; Mismatches 211; Indels 213; Gaps 35;
QY 34 NNYNGVAEVT - - - - -ERQAA - - - - -NGQIGVDGKIISANTTSGSTNOESSA - - - - -TNNT 80

Db 66 NNUNTRVQVNTASSMPNTNPKQATNNSKILVAPKLNQASSPNEGITPKQASIPYNTV 125
QY 81 ENAVNESKNTNTNNAVNE - - - - -NKNTNNT - - - - -ENAVNENKNTNTN 124
Db 126 DN - - - - -KNTFKNSSINNEAPIPKDTSKTSTSAOTKGSNDNNIPSNNTSTNTSKN 179
QY 125 D - - - - -NSOLKIT - - - - -NNEQPSAATOANLKLNPQAQAKAVONAKIDAGSLTDD 169
Db 180 ENPSNDITKTEPAPANAPKDPNNQSDSALAKNALSNNLADSSQTSKVTSN - - - - - 235
QY 170 QINELKNINFSKSAEKGAKLTFKLEGIGNAIVKQDPQYAIPTYSNAKEIK - - - - -NMPAT 224
Db 236 - - - - -NDAPKVNITSTDKKASLNNDSDQD - - - - -WVTKDGKYY - - - - -YVNGVQKGFQFSINKSIY 288
QY 225 YTVD - - - - -AOTGKAHLVDWDSWPQDPVTGYVSNYMGYQLVIAMGIPNPSPTGDNHIYLL 281
Db 289 YFNDGSMQTG - - - - -W - - - - -LKYSNSYFFDASGYML - - - - -TGLQNINGT 326
QY 282 YNKYGNDFSHWRNAGSIFGKTETNVEQWSGSAIVNDDGTIOL - - - - -PFTSNDTSDYKLNQ 340
Db 327 Y - - - - -YGFND - - - - -DGKLLTGLQAIN - - - - -NNYFFNNDGVMTGWTCHTNDSKYFFDNG - - - - - 374
QY 341 RLATATLNLNVDDNG - - - - -VSIKSDVNYQVLFEG - - - - -DGPHY - - - - - 375
Db 375 VMQTLGVHNNKYFGNDGKLLTGLQINNYTYFFDSNGVMQTDWITIDGSKYFFSVNG 434
QY 376 - - - - -QT - - - - -YEQFANGKRENDYCYLRDPHVQLENGDRVLYFEAN - - - - - 414
Db 435 VMQGTIYISGYGFA - - - - -NDGKLLTG - - - - -LOVINGNSY - - - - -YFDTNGIRLVSRWITI 484
QY 415 TGVEDYQSDQIY - - - - -NWANYGGDDAFNIKSSKILLNNKKDRELALGALGILKLTNNQ 472
Db 485 DGKDYFENQDGLTDNWINYDGKYYFISG - - - - - - - - - - -VKQTGLQ 521
QY 473 SKPKVEEYSPVSLTLMACDEVXXKLGDKYLYFSVTRVSRGSRELTAKNTIVGDN - - - - -VA 531
Db 522 N - - - - - - - - - - -IDGNTYFFDSGIMQTLQKIDGK - - - - -TYFEGDNGIR 556
QY 532 MIGVYSDSLMGKYPKPLNNSGVLTASVPANWRTATYSYAVPVAGHPDOVLITSYM 590
Db 557 QIGWITYQ - - - - -NNKY - - - - -FNSDGSMTDLKIYSTSPYNYHYQYGFDFDNDGKLLTGLQTIK 613
RESULT 15
O86489 PRELIMINARY; PRT; 1166 AA.
AC O86489
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SDR E PROTEIN.
GN SDR E.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEWMAN;
RX MEDLINE=99098700; PubMed=9884231;
RA Josefsson E., McCrea K., Ni Eidhin D., O'Connell D., Cox J., Hook M.,
RA Foster T.J.;
RT "Three new members of the serine-aspartate repeat protein multigene
family of Staphylococcus aureus."
RL Microbiology 144:3387-3395(1998).
DR EMBL; AJ005647; CAA0652.1;
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 1166 AA; 126348 MW; 750A7B0135287D4A CRC64;

```

Query Match      4.8%; Score 199.5; DB 2; Length 1166;
Best Local Similarity 22.0%; Pred. No. 0.051;
Matches 166; Conservative 99; Mismatches 312; Indels 177; Gaps 39;

Qy 12 TLVSNILMGVW-----TAHADQ-----VSNYNGVAEYNT 44
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 29 TVGTASILVGTTLIFLGNQAKAAENTSTENAKQDATTSDNKEVSVETENNSTTENS 88
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 45 EROANGQIGVGGKIISANSNTTSGTSSQESSATNTTNAVYVNESKNTNTTNAVYVNEKN 104
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 89 TNPKEKNTDSQPEAKKESSTSSQKQNNVTATTETKPNONIEK-----ENVKPSDKT 143
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 105 TNNTENAVYVNEKNNTNTENDNSOLKLTNNQPSAAQANLKLNPQAAKAVQN----- 158
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 144 ATEDTSVILEEKAPNNTNNDVTKPSTS--EPS-TSEIQTKPTTPOESTNIENSQPOPT 200
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 159 -AKIDAGSLTDQINELNKNFNSKSAEKGAKLTFKDL-EGIGNAIVKQDPOYAIPIYNAK 216
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 201 PSKYD--NOVTDATNPKEPVNSKEELKNPEKULVNRDNDSTHSTKPVATAPTSVAP 258
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 217 EIKNMPATYTYDAQTGKMAHLVDVND-----SWPVQDPVTGYVSNYMGYQLVIAMMGP 269
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 259 KRVNAKMRFAV-AQPAVASNNVNDLIKTKQTKVGGKDNVAAAHGDKDI-----EYD 312
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 270 NSPTGDNHIIYLLYKNYGDN-DFSHRNA--GSIFGTETNVFQEWGSAIYN---DDGTI 323
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 313 TEFTIDNKKV-----KGDWTMTINTDKNVIPSDLTDKNDPIDITDPSGEVIAKGTFRATK 367
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 324 QLEFTSNDTSD-----YKLNQDQ-----LATA-----TLNINVDNNGVS 357
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 368 QITYFTDYVDKYEDIKSRULTLSYIDKKVVPNETSLNLPATAGKETSQNVTVDYQDPM 427
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 358 IKSVDNYQVLFEGDGFHYQYEQ--FAN--GKDREN-----DDYCLDRPHVVOLE 403
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 428 VHGSNIIQSIFTKLDEKQTEQQIYVNPPLKKSATNTKVDIAGSQVDY-----GNIKIG 482
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 404 NGDRYLVEANTGTEDYQ--SDQO-----IYNWANYGGDDAFNIKSFKLLNNKK--DR 453
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 483 NGS--TIIDQNTKIVKVNSDQOLPQSNRIYDFSQYE-----DVTQF-----DNKKSFSN 532
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 454 ELAAGLANGALGILKLTNNQSKPKVEEYISPLV-STLMACDEVXXKLGDKYLFVSVTVSR 512
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 533 NVATLDFGDI-----NSAYIIVKVSXYTPTSDGELDIAQGTSMRTTDKYGYNYA---- 582
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 513 GSDRELTAKNVTIYGDNVA-----MIG-YV-----SDSLMG---KYKPLNNSGVVLT 556
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 583 GYSNFIVTSNDTGGGDTGVKPEEKLYKIGDYVWEDVDKDGVOGTDSEKPMAN--VLVTL 640
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 557 SVP-----ANWRATYSYVAVPVAGHPDQVLI-----TSYMSNKDFASGEGNYATWAPSFL 607
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 641 TYPDGTTSKSVRTDANGHYERGGLDGTYYVVFETPTGYLPTKVNGTTDGEKDSNGSSVT 700
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 608 VOINPD-----DPTTVLARATNOGDWYDDSSRN 636
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 701 VKINGKDDMSLDGTGYKEPKYNLGDYVWEDTNKD 734

```

Search completed: September 26, 2002, 18:28:01
Job time: 260 sec